**DOCTORAL THESIS**

# Investigating the Impact of Focusing on Academic Vocabulary Using Multiple Assessment Measures

Alothman, Khalid

*Award date:*
2014

*Awarding institution:*
University of Roehampton

# Investigating the Impact of Focusing on Academic Vocabulary

# Using Multiple Assessment Measures

Khalid Alothman

A thesis submitted in partial fulfillment of the requirements for the degree of PhD

CENTRE OF LANGUAGE ASSESSMENT RESEARCH,
UNIVERSITY OF ROEHAMPTON,

2014

CENTRE OF LANGUAGE ASSESSMENT RESEARCH,
UNIVERSITY OF ROEHAMPTON

Doctor of Philosophy

# Investigation the Impact of Focusing on Academic Vocabulary

# Using Multiple Assessment Measures

By Khalid Alothman

## ABSTRACT

The academic word list (AWL) is essential for EFL learners wishing to study at university level in English speaking universities. This list, together with the 2000 most common English words, is vital for learners' comprehension of academic texts (Nation, 2006). However, the current practice of many presessional courses in the UK and English foundation courses elsewhere, e.g. Saudi Arabia, does not include the explicit teaching of the AWL within the curriculum, even though research has shown that direct teaching leads to higher pickup rates of the targeted words than implicit teaching (as in Sonbul and Schmitt, 2009). Academic vocabulary is mostly taught within other language input such as academic reading and listening, or made available for students for independent study. For this thesis, three studies have been conducted; the first investigated how much of the AWL is learned on typical presessional courses in an English speaking country - the UK. Two universities hosted this study, accommodating 103 participants in total. The second study investigated how much of the AWL is learned among students receiving two different methods of vocabulary teaching. This study took place at the Preparatory Year (PY) at KSU Riyadh, Saudi Arabia (the largest EFL project of its kind). Unlike most previous work, this study has used multiple measures to assess three levels of vocabulary knowledge: meaning recognition, written form recall aided by initial letters, and vocabulary richness in free-writing tasks. In the third study, learners and tutors were surveyed to state their beliefs, practices and experience in order to record their views concerning many issues regarding vocabulary teaching to advanced learners. The results show that EFL learners at this advanced stage of language study have a poor ability to produce academic words, regardless of their high performance in recognizing the AWL.

Their knowledge of the academic words was found to correlate with general vocabulary size and their ability to write academic words in free writing tasks. Furthermore, learners exposed to direct enhancement of the academic words during the Preparatory Year PY outperformed learners exposed to regular academic teaching in the course. Finally, regarding beliefs and feedback concerning vocabulary teaching, it was found that advanced learners appreciate direct lexical instruction more than implicit approaches, contradicting many common perceptions that advanced learners prefer implicit vocabulary learning. The results show further interesting variations in vocabulary gained and feedback regarding vocabulary teaching across the different groups identified. The implications of these findings for teaching and learning as well as assessment in the academic context are presented and discussed.

# AKNOWLEDGMENTS

''For, were you to count the favours of Allah, you will not be able to count them. Surely Allah is Ever Forgiving, Most Merciful'' (Nahl: 18)

All praise and thanks are due to Allah the Almighty for granting me His endless grace and bounties, and who has blessed me with guidance and wisdom to complete this research. Without His aid and mercy, this thesis would never have seen the light of day.

I am deeply grateful to Professor Barry O'Sullivan, my academic supervisor, for his support, comments and invaluable feedback during all the stages of writing my PhD thesis. His advice, assistance and suggestions were always invaluable and inspiring. I would like also to thank my co-supervisor Dr. Eva Eppler for her useful feedback and support. Very special gratitude goes to Dr Mark Garner, director of CLARe, and to my internal and external examiners. My sincere appreciations also go to the universities hosting my experiments and investigations, my fellow teachers and colleagues.

I am truly indebted to my father and mother for their endless support, advice, encouragements and prayers. Their words of concern for me and my family were always helpful. I also would like to extend my appreciations to my sister Ohoud, Ameera, Hossah and brothers Badr, Fahd, Mohammad, Abdulrahman, Ibrahim, Tamim and Abdulaziz for their prayers, encouragement and support.

A very warm gratitude to my beloved family; my dearest wife Noura, and my son Abdullah, who made my life in the UK easy, enjoyable and pleasant. I greatly appreciate their love, care, understanding and encouragement.

I would like to acknowledge the sponsor of my studies; the Ministry of Higher Education, for granting me this scholarship, and a special thanks to the Saudi Cultural Bureau in London for their support and care during my stay in England.

I am very grateful to all members of the Saudi Society in Southampton, whose friendship has given me and my family great pleasure. Their presence made our staying in the UK enjoyable and unforgettable.

Lastly, I would like to thank all my relatives and friends overseas for their prayers, remembrance and concerns. A very big 'thank you' to you all.

# Table of Contents

## List of Tables

## List of Figures

## List of acronyms used in thesis.

**AWL**: The Academic Word List.

**EAS, EAP**: English for Academic Study/Purpose

**ESP**: English for Special Purpose

**ELT**: English Language Teaching

**GSL**: The General Service List of Acronyms

**IELTS**: International English Language Testing System

**Lemma**: the head word and its most frequent inflections that do not change the part of speech (play, plays, playing not player).

**Presessional course**; short academic English course before the entry of graduate or postgraduate studies.

**PY**: Preparatory Year (English and Math foundation year prior to academic entry study)

**SLA**: Second Language Acquisition

**TOEFL**: Test of English as a Foreign Language

**Tokens**: all words that appear in texts.

**Type**: play and player are different word types

**Word family**: manager, management and manage are members of the same word family.

# 1. Chapter One

# Introduction to the thesis

## 1.1. Overview of Chapter One

This chapter starts with an introduction to the paper, examining some of the issues related to the importance of vocabulary in language teaching. It then discusses the status of lexical teaching research in language research in general, indicating some research gaps and proposes the need for this particular research, stating some background. The nest section looks at motivation and context of the study, followed by a discussion of the significance of the thesis. The chapter concludes with an outline of the anticipated hypotheses and research queries addressed in the thesis, as well as summing up the organization of the thesis.

## 1.2. Introduction.

"Without grammar little can be conveyed; without vocabulary nothing can be conveyed".

(David Wilkins, 1972: 111)

It has been a habit of many fellow researchers of vocabulary to introduce their work with this quotation in order to mark the importance of vocabulary as the quotation sums up most of the literature that has discussed the weight of vocabulary against other language skills for second language learners (see for example Choudhury, 2010; Milton, Wade, & Hopkins, 2010; Schmitt, 2010; Srichamnong, 2008; das Neves Seesink, 2007). Sinclair (1996) adds ‘’a lexical mistake often causes misunderstanding, while a grammar mistake rarely does’’ (*cf* Lewis 1997:16). Hence, Meara, (1996: 35) has argued that "lexical competence is at the heart of communicative competence." Without sufficient vocabulary, learners may struggle with their written and spoken communication. Some

researchers (Adolphs & Schmitt, 2003; Hirsh & Nation, 1992; Laufer, 1992) believe that learners need to know at least 3000-5000 word families, or even 8000 words as suggested by Nation (2006), in order to read comfortably and involve meaningfully in spoken communication at an advanced level.

Unfortunately, as Schmitt (2008a) reports, many students at the academic level fail to reach such a threshold, and may not even reach the 2000 general service list. Even high level second language learners (L2) high school graduates may not necessarily master the academic words (AWL) or the English for special purpose (ESP) vocabulary that appears frequently at graduate levels. Furthermore, even those who can recognize a threshold of 2000 or 3000 words may not be able to produce them perfectly or freely in writing tasks. Therefore, teachers and learners have real concerns regarding how to efficiently reach the proposed required level of vocabulary size in order to overcome any language learning difficulty, both receptively and productively, at the academic level.

There are quite a few studies that have investigated the vocabulary level of students and explored how it correlates with their language proficiency (Stæhr, 2008). However, although the vocabulary size of learners had been determined, there is little known regarding what the specific vocabulary level of a student can tell about his/her ability to use the words he/she knows in a free or controlled writing task.

Thus, the present thesis investigates how relatively advanced students at the academic level best learn the general academic vocabulary, and how far this affects their other language skills (in this case writing). It also explores how far multiple measures can be useful to determine the learners' knowledge of vocabulary. The study moves on to

ascertain the effectiveness of the methods of vocabulary learning used, both from the teachers' and the learners' prospective.

### 1.3. The need for the research.

As we have seen above, vocabulary has been regarded as the most basic language component. Without an extensive size of vocabulary, even advanced EFL learners may not perform well in their academic spoken or written communications. Regrettably, as Schmitt (2008a) reports, most EFL high school graduates have only reached, or sometimes even fall short of, the 2000 word level (the General Service List GSL compiled by West (1953) see 2.3.1). With such a little vocabulary size, learners may be challenged when reading or communicating at the academic level. Even with their mastery of the most frequent 2000 words, they may not be able to read academic texts comfortably. Therefore, the learning of the AWL is essential, or at least, should be a minimum objective for students aspiring to graduate level study. As research has suggested, the academic word list (570 headwords), together with the most frequent GSL, make up to nine in every ten words that appear in English texts (Coxhead and Nation (2001), see 2.2.2.). Of course, as mentioned above, researchers have recommended even higher vocabulary sizes for optimum comprehension of texts as the challenging figure suggested earlier by Nation (2006).

In Saudi Arabia (SA), the host of the second study, the vocabulary size of university students ranged between from 1000 and 3000 words, depending on the type of the test used, and level and major of students. AlQahtani (2005) gave 85 students (majoring in English in their second year of study) a receptive test, and found that they knew 3000 words. With similar subjects, Alfraidan, (2010) found that the students' productive

vocabulary (words that they could produce) was 1,135 words. Using a yes/no question format, AlSaif (2011) found that students at the junior level of the English department appeared to know on average of 2,452 words. Studies involving general SA high school graduates have found that learners have poor vocabulary levels, scoring below 1000 words (AlAkloby, 2001; AlNujaidi, 2003; AlSaif, 2011). AlQahtani's (2005) data, however, shows that high school graduates knew an average of 2327 words. Nonetheless, AlSaif (2011) has some concerns regarding Alqahtani's method for working out the learners' size (level) of vocabulary, as the known words from the *10,000 word level items* should not contribute greatly to the overall score of students. i.e. the knowledge of few words at the 10k test items should not contribute to a higher overall score of the test (the VLT is reviewed in see sections 2.4.4 and 3.4.1).

At any rate, the data shows that even higher achieving students who know over 2300 words would need the AWL for their academic study in order to reach the higher comprehension levels needed to read advanced texts. As argued by Alothman (2011), such students are considered among educators to be upper-intermediate and/or advanced EFL learners, and therefore, there is a little focus on vocabulary (see 2.3.).

Thus, the current thesis aims to provide a clearer idea of what constitutes best practice regarding the teaching of academic vocabulary (i.e. explicit or implicit) to intermediate and advanced students, and how far it benefits their language skills.

There might indeed be a great need for more research into the vocabulary size of students, since some previous studies only provide a broad estimate of the vocabulary size level (see Webb, 2008). Part of this may have resulted in the variety of results revealed in the studies reported earlier about Saudi participants' vocabulary level size.

This thesis also aims to gain more information about the students' knowledge of vocabulary using multiple measures to assess vocabulary knowledge.

## 1.4. Motivation and context of the study.

There is no "right" or "best" way to teach vocabulary, as this depends on many factors e.g. level of students and targeted words, and time allocated for learning (Schmitt, 2008b). Hence, a significant motivation for this research is to investigate which method of vocabulary teaching is more efficient, leads to greater gains and is best applied to advanced level students, especially those studying in intensive and time-limited courses. Outlined below are some more reasons for this research:

1- As argued by Alothman (2011), many English for Academic Study (EAS) courses pay little attention to teaching vocabulary to advanced students as it was always thought that learners at such a stage of proficiency, are more likely and capable of learning new vocabulary from reading and listening. Thus, this research marks an attempt to gain concrete evidence about which method of vocabulary teaching (either incidental or explicit) leads to higher pick up rates of new vocabulary with advanced students.

2- a) The research above has pointed out the importance of the AWL and its frequent appearance in English texts. However, from a closer and more detailed analysis of the data of (AlQahtani, 2005; AlSaif, 2011; Alfraidan, 2010), it could be argued that even advanced EFL learners majoring in English who are in their second academic year have a lower academic vocabulary level than recommended by research. These figures revealed that students know the equivalent of less than one fifth of the academic vocabulary. This means that

they may have problems with learning language, both receptively and productively, with such a proportion.

b) What can be drawn from this is that although students in these contexts have been studying an English major degree for a year or more (sometimes plus the intensive English course), the figures show lower pick up rates of academic vocabulary than expected from exposure to such an academic English environment.

3- There are several studies that have involved Saudi students investigating the issue of extensive reading and its impact on vocabulary growth (AlHomoud, 2007; AlNujaidi, 2004; AlRajhi, 2004). However, these studies have targeted students with low vocabulary sizes and, therefore, have used graded readers that might not be suitable for more advanced students. The texts of graded readers mostly contained a range below 1800 words and, therefore, might not contain the AWL (Horst, 2005). Moreover, even those studies, e.g. AlHomoud (2007), who has used textbook reading passages for intensive reading, might not cover the AWL either, since textbooks materials do not always address the issue of vocabulary frequency (Milton, 2009). Therefore, there is a great need to investigate which method is best applied to achieve better results of vocabulary growth for more advanced students.

4- As argued earlier, there is also a need to investigate other aspects of word knowledge rather than just a broader estimate of the vocabulary size. This should help to explore what learners of a certain vocabulary size are capable of regarding other aspects of word knowledge (namely free and controlled

productive knowledge). Besides, using multiple measures is likely to provide clearer results regarding the learning outcomes of the students.

5- In Saudi Arabia, there is a growing number of universities that have started to run *foundation*, *academic* or *Preparatory Year (PY)* programmes and include all graduate students, not just those who get into the schools where English is the main language of instruction. Some of the PY programmes, such as the programme hosting this study, could accommodate up to 8000 male and female students. There has been a particular interest amongst PY administrations in finding the best means for teaching all English skills, and in particular, what contributes most to the development of their writing and reading skills. The present study can feed into such programmes by providing some investigations to ascertain which effective methods need to be considered, especially concerning those students at the advanced levels of these courses.

6- Related to this, the results of this study would be of great use to some specific graduate schools which have plans to teach some English for Special Purpose (ESP) words to their potential students.

7- Several studies, some reviewed here, have looked at the multiple measure of vocabulary because of the complex nature of word knowledge. This study will contribute to the issue as it uses multiple measures that test word knowledge. Interestingly, this study would have further input regarding the issue since different word learning approaches were involved.

8- Studies show that L2 vocabulary retention is higher for students who complete written vocabulary activities after a reading task than for students who complete

another reading assignment, (with the same target vocabulary), after the reading task (Folse, 2006; Hulstijn & Laufer, 2001; Paribakht & Wesche, 1999; Zimmerman, 1997a). This supports the benefits of the *involvement hypothesis* on vocabulary learning (Hulstijn & Laufer, 2001; Schmitt, 2008a). Looking at this from a different perspective, the present study considers how far presenting the general AWL to learners is actually allowing earlier noticing of vocabulary, and can this result in deeper processing while performing general academic skills.

9- Nation (2006) suggested that a vocabulary size of 8000-9000 words should be recommended for L2 learners aiming to study at the academic level, though some researchers have concerns about this as it may seem very challenging (Laufer & Ravenhorst-Kalovski, 2010; see more in section 2.2.2.). Laufer and Rozovski-Roitblat (2011) suggested two figures; an optimal 8000 words level for higher lexical coverage of texts, and a minimal level of 4000-5000 words for a modest lexical coverage (see 2.2.3.). From my L2 perspective, the 8000 word level appears to be very optimistic. Besides, many of the studies reported earlier which relate to Saudi learners suggest that even high achievers are unlikely to reach the 8000 word level, and yet gain admission to academic study. Therefore, since the present study uses a vocabulary size test parallel to the learners' performance in the AWL, the potential results should be insightful regarding which figure, 8000-9000 words or 4000-5000, is the appropriate objective for L2 learners.

10- This study uses a vocabulary size test to investigate the vocabulary growth of learners studying on comprehensive English courses. It speculates about the rate

of vocabulary gain within a single academic term so that it can recommend a goal that suits such two-term intensive programmes.

11- Alemi and Tayebi (2011) could not find any significant difference between incidental and intentional learning of vocabulary. In their discussion of limitations, they suspected that some attention must have been involved in the supposed incidental vocabulary (more discussion and remarks in Section 2.3.2.1.). The current study, however, is expected to have clearer outcomes as it compares the results of two different groups exposed to two methods of teaching, and involving a larger vocabulary target size. Alemi and Tayebi (2011) targeted only 12 words.

12- Llach and Gallego (2009) examined the relationship between receptive vocabulary size and written skills of Spanish primary school learners. They found a correlation that was not very high (.542 for or the 1k of the VLT and .503 for the 2k), even though it was significant. However, a study involving non-Latin participants, such as the present study that looks at only Arabic students, may give great insights into the issue.

## 1.5. The importance of the thesis.

The debate over the best approach to vocabulary teaching is ongoing. As was mentioned earlier, Schmitt (2008b) stated that no method would have an advantage over the next as "each *situation* will depend on different factors such as the type of student, the words which are targeted, the school system and curriculum." The importance of this work regarding this issue lies on its implications for one of these "*situations*" which is; 'the learning of the academic vocabulary by advanced EFL learners in intensive English

courses'. Through comparative work, this paper will provide, with concrete evidence, some useful insights regarding the best way to deliver vocabulary teaching to advanced EFL learners studying on academic courses that apply: *none* (only incidental learning of vocabulary through exposure), *explicit vocabulary teaching,* and *studying vocabulary independently*. Although explicit teaching of specific vocabulary results in greater and faster gains than incidental learning from exposure only in the classroom setting (Sonbul & Schmitt, 2009), there is still a great need to investigate how a larger corpus of vocabulary like the AWL can be acquired in a longer scale study. Of interest also is the fact research involving Saudi students (as seen above in AlNujaidi, 2004; AlQahtani, 2005; AlSaif, 2011) shows poorer gains of the academic vocabulary than expected from learners studying academic English for quite some time.

This paper aims to ascertain which approach leads to greater gains of the AWL, and how far the AWL benefits high achiever learners' language skills (writing mainly). It attempts to explore whether presenting the AWL to advanced learners, rather than just learning the items contained in it incidentally, actually makes a difference, and if so, to what degree. Furthermore, if advanced learners already show higher levels of vocabulary (receptive or recognition) knowledge, then, what can we learn about their needs regarding the *depth* of knowledge? It is additionally anticipated that the findings of this study will be useful for language teachers in general and intensive course coordinators in particular in outlining a structured approach to vocabulary teaching, to know what is feasible within a limited time, and to plan more effective materials for their courses.

The results regarding learners' beliefs and attitude may help material developers in preparing and developing exercises that improve autonomous learning.

## 1.6. Summary of organization

Following the introductory chapter (Chapter 1), Chapter 2 will focus on an account of the relevant literature. It will provide theoretical background to a number of issues, including an introduction to the importance of vocabulary on language competence, and how high frequency words contribute to texts and, therefore, to learners' comprehension. It then discusses how vocabulary is learned; this is broken down into investigating the concept of ward knowledge and what does word knowledge involve. The variation between vocabulary depth and breadth of knowledge is discussed. An argument about approaches to vocabulary teaching will then follow. The key studies that have involved vocabulary learning through two main approaches, incidental and explicit will be reviewed. A concluding argument about best practice regarding the method of teaching, especially with advanced learners, will be presented.

The chapter ends by discussing the issue of vocabulary testing. It begins with the importance of vocabulary testing, then states why multiple measure are crucial in vocabulary research. It reviews some of the main measures of vocabulary size, and then shows why the particular research tools used were chosen for the present study.

Chapter 3 describes briefly the methodology of the studies of this thesis. It first gives a general rational for the thesis and what each study is expected to contribute, describing the general design and type of methodology used in these studies, and how these differ in terms of their aims and research questions. The chapter moves on to describe the instruments used in the studies and how they were adjusted to serve each individual study. It continues by describing the participants and the background of the studies, and then follows with a brief description for procedures of the data analysis.

Chapter 4 describes *the first study*. It gives some background about the contexts of study, and then states the aims and objectives, as well as describing the participants of the study, the instruments used and the procedure. It then reports how the data is be analyzed. The results of each research question are then presented and the findings discussed.

Chapter 5 is the *second study*, in which the target research is described in detail and the results presented to the reader. This is followed by Chapter 6, which is focused on the *teachers and learners' beliefs and feedback*. These chapters follow the same organization of the earlier chapter (Chapter 4). Chapter 7 summarizes the general results of the studies and elaborates with some discussions of the major findings. It concludes by proposing some research and pedagogical implications of the thesis. Chapter 8 concludes the thesis, presenting some limitations of the studies and suggests some recommendations for future research.

# 2. Chapter two

# Literature review

## 2.1. Overview of Chapter Two

It is important to discuss a number of related issues to establish a better understanding of the main two themes of this thesis; vocabulary learning and vocabulary assessment. It is, first, essential to shed some light on the importance of vocabulary for language learners and its significant impact on language skills. This will include exploring the possible number of words that might be needed, and what words frequently appear in spoken and written English texts. Further explanation will be provided in terms of defining exactly what constitutes a word, and what word knowledge entails. Here, the discussion will specify the receptive and the productive knowledge of vocabulary in more detail. Breadth and depth of vocabulary knowledge will be covered also in the discussion in relation to these types of knowledge in terms of teaching and assessment. After determining the aspects of word knowledge, it is equally important to discuss how vocabulary is *learned* and how it is *assessed*. First, an overview of the approaches to vocabulary teaching will be provided to set up the rationale for this study to determine three issues; what is the best approach to vocabulary teaching, how much is acquired incidentally from exposure and how much is learned from direct teaching.

This literature review concludes with a comparison between the two approaches of vocabulary teaching and suggests the best approach to apply in which learning context. The second half of the chapter is about vocabulary testing. The first issue to be addressed here is the importance of vocabulary testing and its proposed position in terms of overall testing of proficiency. A review and discussion is provided about the following issues; vocabulary test formats, vocabulary size tests, and validity of tests. The chapter will end

by linking the previously discussed issues (aspects of word knowledge, how vocabulary is learned and how it is assessed at present) and will conclude by addressing the need for multiple measures and what this tells about students' vocabulary learning.

## 2.2. The importance of vocabulary.

Vocabulary is crucially important for any language use. Without words, spoken and written communication would not be possible. Unfortunately, until the 1980s, grammar was given dominance over vocabulary in the area of language teaching research. Fortunately, as scholars report, there is now a very substantial body of research available on vocabulary learning (Laufer, 2009; Nation, 2001; Read, 2000; Schmitt, 2008a) Recent research on language learning has started to give vocabulary a higher status when compared with other language skills. Zimmerman, (1997b: 5), for instance, comments such as "vocabulary is central to language and of critical importance to the typical language learner." Gass and Selinker (2001: 372, *cf* das Neves Seesink, 2007) indicate that "…there are numerous reasons for believing that lexis is important in second language acquisition, in fact, the lexicon may be the most important language component for learners." These views are now shared by many researchers and linguists (Folse, 2011; Laufer, 2009; Meara, 1995; Nation, 2001; Schmitt, 2000, 2008a, 2010a), and this has translated into many works and papers involving vocabulary in language education (Coady, 1993; Coady & Huckin, 1997; Laufer, 2009; Milton, 2009). Research has looked at many areas regarding vocabulary learning such as; ways and methods of teaching vocabulary (Folse, 2004; Nation, 1990; Schmitt, 2008a, 2008b), techniques to deliver vocabulary (Oxford & Crookall, 1990), word size, word frequency (Coxhead & Nation, 2001; Laufer, Elder, Hill, & Congdon, 2004; Laufer & Ravenhorst-Kalovski, 2010;

Nation, 2006), vocabulary testing and ways of assessment (Daller, Milton, & Treffers-Daller, 2007; Milton, 2009; Read, 2007; N. Schmitt, 2010b) and others. Word frequency and text coverage, in this researcher's view, have played and will continue to play a central role in the direction of the research of second language (L2) learning, teaching and assessment. To illustrate, in the research into word frequency, learning and assessing vocabulary should be the focus. Moreover, more tests could be developed to target frequent words from different disciplines. Furthermore, material writers will continue to develop textbooks and learning materials that target frequent words and words for special purpose. The following sections discuss this in more details (see 2.2.2. and 2.2.3.)

### 2.2.1. Vocabulary and language competence

Empirical research shows a direct correlation between the learners' vocabulary and their language skills (Laufer, 1992, 1997; Llach & Gallego, 2009; Mehrpour & Rahimi, 2010; Pretorius, 2000; Qian, 2002; Stæhr, 2008). Researchers have found that the size of vocabulary correlates positively with learners' skills i.e. reading, writing, listening and speaking. Reading has received much focus in the research as it is perceived as the most important skill for academic success (Laufer, 1992; Pretorius, 2000). Vocabulary is assumed to have an important influence and is related to comprehension (Laufer, 1997; Llach & Gallego, 2009; Mehrpour & Rahimi, 2010). Researchers have found that vocabulary knowledge and reading performance strongly correlate, for example: Laufer (1992) found a correlation figure of .50–.75, and in Qian (2002) it was .73–.77. Vocabulary size has been found to predict as much as 72% of the variance in reading (Stæhr, 2008). Qian (2002) believes that scores on *depth* of vocabulary knowledge (how well students recognize the different uses/aspects of individual words) can make a unique

contribution to the prediction of reading comprehension levels. Tannenbaum, Torgesen, and Wagner (2006) found that breadth (how many words are recognized by a learner) has a stronger relationship to reading comprehension than depth/fluency; however, the two dimensions of word knowledge have significant overlapping variance that contribute to the prediction of reading comprehension level.

The impact of vocabulary on writing is equally important (see the tables and discussion below). An obvious and easy way to observe this is by reading and comparing written performances of L2 and native speaker students. Native speaker students mostly tend to use more sophisticated words than L2 writers (Laufer & Nation, 1995). Morris and Cobb (2004) found that academic words appear less in free writing tasks of English non-native graduates than English graduates. AWL researchers have highlighted the importance of vocabulary on writing in their work (Astika, 1993 *cf* (Stæhr, 2008); (Daller & Phelan, 2007; Laufer & Nation, 1995; Lee, 2003; Llach & Gallego, 2009). Stæhr (2008) and Milton et al., (2010) showed a correlation of (0.73), and (0.761) between vcabulary and writing. However, Stæhr (2008) admits that the relation between vocabulary size and ratings of written compositions is a complex issue. He argues:

> '*the studies that found the relationship between the two variables cannot necessarily be taken as direct evidence that learners with large vocabularies are better at writing*' p. 141

This suggests the need to collect more information about the students' level and for detailed investigation into how vocabulary contributes to writing.

The link between speaking, listening and vocabulary has been given less attention, based on the literature being reviewed. Nonetheless, some empirical research has shown the correlation between vocabulary and speaking (Borer, 2007; Hincks, 2003; Joe, 1998) and listening competence (Ching-Shyang chang, 2007; Mehrpour & Rahimi, 2010; Stæhr, 2009). Stæhr (2008), and Milton et al. (2010), explored the correlation between the lexical knowledge of the L2 learners across multi language skills. Stæhr (2008) investigated the relationship between the vocabulary size of 88 L2 students and their scores in reading, writing and listening examinations. The relationship is summarized in the table below:

**Table 2.1 Spearman correlations between scores of vocabulary size and reading, listening and writing scores (Stæhr, 2008, p. 144)**

| | Listening | Reading | Writing |
|---|---|---|---|
| Vocabulary size | 0.69* | 0.83* | 0.73* |

* Correlation is significant at the 0.01 level

As Stæhr commented earlier regarding writing, participants' compositions were judged against a holistic rating scale that mainly related to the quality of the language, content and organization of writing. Lexis is included within the rating criteria of linguistic aspects of the writing, which may therefore be considered an indirect link. Thus, investigating the link between vocabulary and writing using direct assessments is recommended.

It is important to note that Stæhr used receptive tests in his investigations for the three skills above. It might be recommended, however, to include a productive measure when investigating writing. The table above shows that reading correlates with vocabulary higher than listening. The correlation of writing was expected to be slightly lower than

the figure shown above in the table due to the use of receptive measure used in the study, as Stæhr pointed out.

Milton et al. (2010) included speaking in their study and found a modest correlation between students' IELTS test scores and the two tests they used to measure vocabulary, i.e. the orthographic vocabulary size test *X-lex* (Meara & Milton, 2003) and the phonological counterpart *A-Lex* (Milton & Hopkins, 2005) (see further description of the tests in section 2.4.4). The Table 2.2 below summarizes the results.

**Table 2.2 Spearman correlations between vocabulary size scores and IELTS scores (Milton et al., 2010, p. 91)**

| | A-lex | Read | Listen | Write | Speak | Overall |
|---|---|---|---|---|---|---|
| X-lex | 0.456* | 0.699** | 0.479** | 0.761** | 0.347 | 0.683** |
| A-lex | | 0.217 | 0.676** | 0.441* | 0.713** | 0.546** |

** Correlation is significant at the 0.01 level
* Correlation is significant at the 0.05 level

The table shows that the results of learners' orthographic vocabulary size correlate strongly with their reading and writing skills, and much more moderately (at below .5) with listening and speaking skills; however the correlation with listening is still significant. In contrast, the phonological vocabulary size correlates more significantly with listening and speaking, than with writing and reading. Not surprisingly, as the researchers note, the two correlations suggest that knowledge of the written forms of the words may not be essential in order to perform well in speaking and listening. Besides, the link between speaking and size of vocabulary does not seem to be clear, as speakers with less vocabulary could communicate perfectly (Milton et al., 2010).

By applying two measures to assess vocabulary, Milton et al (2010) seemed to have explained in more detail the link between vocabulary size and overall language skills. For

example, as they noted, their results "would explain why vocabulary size measures in the past have failed to predict oral fluency satisfactorily'' (Milton et al., 2010). However, further investigations within the specific skills, and gathering more information about the learners is recommended in order to obtain more accurate results, and to explore learners' language needs.

What is relevant here is that Milton et al. (2010) found that learners from different backgrounds varied on their performance in the two test types. The Arabic-speaking language group tended to have similar vocabulary size of the *Aural-lex* and *X-lex* tests score. In fact, they were the only language group whose phonological score was slightly higher than their orthographic test. See Table 2.3 below:

**Table 2.3 Mean vocabulary size scores broken down by language group (Milton et al, 2010)**

| | Arabic n=10 | | Chinese n=10 | | Other n=10 | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| X_Lex | 2410 | 322.14 | 3272.22 | 625.55 | 2895 | 1091.49 |
| A_Lex | 2470 | 904.06 | 2394.44 | 658.80 | 2290 | 795.40 |

Milton et al. (2010) however, note that learners with a lower vocabulary size, ranging between 2000-3000 words, might even have a fairly good lexical coverage in academic aural texts, but this knowledge might not be sufficient for handling written texts.

To conclude, it could be argued that:

- Multiple measures are important for extra needs assessment of the learners, especially across multi skill needs.
- Extra information about the individual learners is very useful in explaining the scores, or deciding the assessment tool.

- Although Stæhr (2008, 2009), Milton et al. (2010) and Llach & Gallego (2009) have used a receptive vocabulary test which showed a significant correlation with writing, it is still worth investigating how a productive written measure of vocabulary size correlates with the receptive test rather than with general composition rates. In other words, general composition rates seem to give a broader estimation about academic writing abilities (correct use of language, punctuation, accurate expression, use of sophisticated words etc.). A productive written measure of vocabulary size would result in more accurate results regarding lexis abilities.

### 2.2.2. Vocabulary and coverage of texts and word frequency

Many researchers have responded to Alderson's (1984) seminal article "Reading in a Foreign Language: A Reading Problem, or a Language Problem?" They found evidence that both general reading skills and proficiency in the L2 are important for reading in foreign languages (Grabe, 2004; Perkins, Brutten, & Pohlmann, 1989; Pichette, Segalowitz, & Connors, 2003 *cf* Laufer & Ravenhorst-Kalovski, 2010). Most researchers agree that high achiever language learners function more efficiently with general reading skills (ibid). However, the link between lexis and reading comprehension had not been determined until recently.

According to Laufer's *Research Timeline* of the SLA of vocabulary (2009), Laufer (1989) was one of the earliest researchers to investigate the link between reading comprehension and lexical coverage. By this is meant the link between lexical items (words) known to the reader in the text, and their contribution to overall comprehension. She used learners' self-report of unknown words to calculate the coverage of texts compared to level of comprehension, and found that learners with 95% coverage of

known items of written texts are the most successful in reading comprehension. She suggested that 95% and above is sufficient for the learners' to achieve comprehension (Laufer, 1989). Researchers have continued to work out the percentage and, also, have included the coverage of spoken discourse.

Laufer's (1989) figure has been updated by Hu & Nation (2000) who suggested the percentage to be 98–99%. In their study, Hu and Nation (2000) examined the effects of unknown vocabulary on reading comprehension using four versions of texts rated according to lexical coverage: 80 %, 90%, 95%, and 100%. Their study showed poorer comprehension with learners from the 80% lexical coverage condition. Regarding spoken discourse, although it may vary according to the degree of formality (Adolphs & Schmitt, 2004), Larson and Schmitt (n.d. *cf* Schmitt, 2008) suggest that a coverage of around 90% might be adequate. Their participants showed similar degrees of comprehension from passages with 95% and 100% coverage of known items, and also from passages with 90% coverage, learners showed similar comprehension with a 97.5% coverage. These figures suggest that fairly extensive vocabulary is required for learners to read efficiently and listen with ease in L2 contexts. This begs the question "how large a vocabulary size should English learners possess to reach such coverage?" This question is addressed in the following section.

### 2.2.3. Word frequency lists, and how many words are needed for text coverage?

Studies show that the average educated native speaker of English appears to have a range of vocabulary size of 15,000-20,000 word families (Goulden, Nation, & Read, 1990) Fortunately, learners do not need to master all of those words, as they can function efficiently in language with a much smaller vocabulary size. Researchers have divided all

English vocabulary into groups (or lists) according to frequency of appearance in texts. One of the best descriptions of the words that appear in English texts is Coxhead & Nation (2001); they stipulated four levels of word frequency: high frequency words, academic vocabulary, technical vocabulary, and low frequency words. *High frequency* words are the most frequent 2,000 English words, as they cover almost 80% of the running words (words used) in academic text. The list was compiled by West (1953) who called them the general service vocabulary list (GSL). *Academic vocabulary* is less common in conversations and general texts, but appears very commonly in academic texts. It consists of 570 word families forming the Academic Word List (AWL) (Coxhead, 2000). It covers about 8.5%-10% of the running words (all words used) in academic texts, and therefore, adds extra 10% to the 80% mentioned earlier (this additional 8.5-10% is very significant for text comprehension, as seen in section 2.2.3.). Technical vocabulary appears less commonly depending on the subject area (for more detail, see Chung & Nation, 2003). It could provide coverage of up to 5% of the running words. The remaining proportion of texts (around 5%) consist of low frequency words of English and proper nouns.

Based on their research of comprehension levels, many authors have also analyzed the vocabulary size levels required for comprehension. Laufer (1989) has suggested that at least 3,000 word families (the words: 'running' and 'ran' are family members of run), corresponding to about 4800 words, are required for text comprehension (according to Laufer, the figure of comprehending texts is 95%). Adolphs & Schmitt (2003) found that 2,000 word families (4000 individual words) are sufficient if 95% coverage is required for comprehension. In a further study; Adolphs & Schmitt (2004) examined wider spoken

contexts of CANCODE[1] and suggested that 4,000 word forms are required for above 96% coverage (noting the variation depending on context).

Nation (2006) did a more comprehensive study to explore what vocabulary size is required for unassisted comprehension. He trialed fourteen 1,000 word-family lists created from the British National Corpus BNC and checked what coverage they provided for a variety of spoken and written texts. A summary of his findings is in Table 14

**Table 2.4 Word levels and spoken and written coverage (Nation, 2006: 79).**

| Word levels | Approximate written coverage | Approximate spoken coverage |
|---|---|---|
| 1st 1,000 | 78–81 % | 81–84% |
| 2nd 1,000 | 8–9 % | 5–6 % |
| 3rd 1,000 | 3–5 % | 2–3% |
| 4th–5th 1,000 | 3% | 1.5–3% |
| 6th–9th 1,000 | 2% | 0.75–1% |
| 10th–14th 1,000 | < 1% | 0.5% |
| Proper nouns | 2–4% | 1–1.5% |
| Not in the lists | 1–3% | 1% |

The first thousand most frequent word families would provide a coverage of 78% to 81%, the second thousand adds 8% to 9%, 3000 words add 3-5%. Laufer & Ravenhorst-Kalovski (2010) concluded from this that the findings supported the argument that 95% or a vocabulary level of 3,000 words was required, as had been suggested by Laufer in 1989. However, Laufer and Ravenhorst-Kalovski (2010) felt that Nation's recommendation of 8000 to 9000 for a 98% level of comprehension was 'a very safe suggestion'. They went on to propose a more achievable figure. They proposed two thresholds: an optimal figure, which is the knowledge of 8,000 word families yielding the coverage of 98% (including proper nouns) and a minimal one, which is the knowledge of

[1] The CANCODE stands for Cambridge and Nottingham Corpus of Discourse English ,and it is a 5-million word corpus of spontaneous spoken discourse, which was compiled in the mid 90s.

4,000–5,000 word families resulting in the coverage of 95% (including proper nouns). Their measures were: the levels test, lexical coverage (the newest version of Vocabulary Profile) and reading comprehension in a standardized national test. Laufer and Ravenhorst-Kalovski (2010) were able to link the coverage to reading comprehension with more accurate figures as they combined learners' reading score with the lexical coverage and vocabulary levels test. They have shown concrete evidence of how the percentage of coverage contributes to comprehension. For example, interestingly, they found that the extra 3% coverage that is provided by the third 1000 words, led to an increase of 12.3 points on the reading score (more figures in ibid: 24). Hence, it is no surprise that Hunt & Beglar (2005: 24) argue that "the heart of language comprehension and use is the lexicon."

With respect to the different vocabulary thresholds suggested by Nation (2006) and Laufer and Ravenhorst-Kalovski (2010), the researchers focused on general frequent word bands from 1000 words to 14000 or 20000 words. There was no specific focus on the AWL in their assessments, even though it represents 8-10% of words used in English texts as discussed before. The current thesis investigates the general vocabulary size of advanced EFL learners, and assesses their ability to recognize and produce the AWL. The current study also which the two thresholds suggested above is achievable among advance EFL learners before that start their academic studies.

### 2.2.4. Learners' report on the importance of vocabulary.

There has recently been a substantial increase in language education literature investigating language learning and pedagogy from the students' rather than teachers' perspective. There is a great body of research that explores learners' opinions on needs

analysis, evaluation, techniques of education and beliefs which influence their learning (e.g. Nikitina & Furuoka, 2006: and several other studies that used Horwitz's model, (1988). Learners' feedback is crucial as it informs educators of their understanding and thoughts about language learning.

One of the most well-known models that investigated the beliefs of learners about language learning in general, and widely used in many studies, is that of Horwitz (1988) *Beliefs About Language Learning Inventory* (BALLI). This is a 34-item questionnaire that asks respondents to report their degree of agreement/disagreement with a number of statements related to their language learning. However, regarding vocabulary in particular, there seems to be an absence of studies that report learners' feedback on lexical learning more thoroughly. A few studies have reported on learners' views on vocabulary in their broader investigations of learners' beliefs on language learning in general. For example, in Horwitz's (1988) application of BALLI, item 16 asked three groups of learners learning German, French and Spanish to declare their views on the following statement:

| | |
|---|---|
| ' Learning a foreign language is mostly a matter of learning a lot of new vocabulary words' | 1 = strongly agree, 2 = agree,<br>3 = neither agree nor disagree,<br>4 = disagree, 5 = strongly disagree |

25 to 39 % of the students in each language group agreed that the most important part of learning a language is learning words.

Bernat (2006) confirmed these results by using BALLI with 262 EAP language learners at an Australian university, and compared these results to those of Siebert (2003) who undertook a study in the American context. The results are outlined in the Table below:

**Table 2.5 Item 17 about vocabulary from BALLI results adopted from (Bernat, 2006)**

| Items | | | A | N | D |
|---|---|---|---|---|---|
| 17 | The most important part of learning a foreign language is learning new words. | Aus | 46 | 31 | 23 |
| | | US | 57 | 29 | 13 |

A=collapsed scores for Strongly Agree and Agree; N= Neutral; D=collapsed scores for Disagree and Strongly Disagree.

Again, only a single item about vocabulary was reported. There are further examples of studies that reported comments or results of learners' feedback within a broader investigation declaring the importance of vocabulary for their skills. For example, Leki and Carson (1994) investigated students' perceptions of EAP writing instruction and writing needs after learners entered the university and majored in different disciplines. Their respondents reported, in the open-ended survey, their need for vocabulary instruction. Among the most frequently expressed specific needs from ESL writing courses was vocabulary as at the top of students' demands (38% of the survey).

Zimmerman (1997a) reported that EAP learners ranked in-class vocabulary activities as more useful than learning words from reading. Zimmerman (1997a) comments that learners believe that more attention to vocabulary would be beneficial. Regarding this issue, Krashen (1989) comments that L2 acquirers always report their lack of vocabulary as a major obstacle in language learning. That is why he comments, "they carry dictionaries around with them, not grammar books'' (Krashen, 1989: 440).

## 2.3. Vocabulary Learning.

As stated earlier, vocabulary was a neglected aspect in language learning research (Meara, 1980). Clear evidence of this can be inferred from the extensive research as well as the focus on ELT approaches which target functional linguistic aspects rather than

lexical aspects. For example, as Schmitt (2000: 12) argued that *the Grammar Translation Method* was based on the emphasis of teaching grammar rules explicitly, and "vocabulary was only selected to illustrate those rules." Also, in *Communicative Language Teaching*, the focus is mostly on meaning (and messages) and vocabulary is taught only for language use as "it was assumed that…'it' would take care of itself" (Coady, 1993). The *Direct Method* and the *Audio-Lingual Method* involved teaching vocabulary more explicitly, but still, the focus was mainly on function, and vocabulary was used in more sentence drills; see Alothman (2009) for a deeper discussion on this. However, as indicated above, this is no longer the case (i.e. less interest in vocabulary) as research on vocabulary learning has increased significantly (Laufer, 2009; Read, 2007; Schmitt, 2008).

It could be argued that there were some research areas that contributed positively to 'the shift' in the renewal of interest in vocabulary research. Corpus-based research has greatly influenced SLA vocabulary research. These include researching the frequency of words, investigating the coverage of common or specific words in English texts, compiling ESP vocabulary that appears in specific texts, deciding the percentage of vocabulary levels in relation to learners' language performance and, most importantly, the vocabulary tests that suggest the required sizes for more advanced language competence, and for better and clearer diagnosis of language needs.

These research areas have led to recognition of the significant contribution of vocabulary to the performance of language learner skills (reading, writing, listening and speaking, as discussed earlier in the chapter). For example, in Laufer's review of the SLA vocabulary research '*timeline'* (2009) she argues that, in the late 1980s, she was the first to correlate

lexis to reading comprehension and state that vocabulary is probably the main key link to success in reading.

Another example that was inferred from Laufer's review of the *research timeline* that supports the current author's proposal, is Qian's (1996) challenge to the *focus on message* and the discouragement of the decontextualization of vocabulary learning during the heyday of the *communicative approach*. Qian (1996) compared teaching contextualized vocabulary with decontextualized word lists and found that the decontexualized condition resulted in better retention. These results challenged the assumptions by Krashen (1989) and Long (1991) (i.e. language acquisition via focussing on the message). However, Laufer (2003) reached a similar conclusion.

Word lists have also presented a basic and achievable goal for learners and have supported further investigations of direct teaching. Learning vocabulary through graded reader stories (Nation & Wang, 1999) for beginners was proposed as a response to opponents of adopting extensive reading programmes to extend vocabulary, and offered a suggestion for their practical limitation.

Corpus-based studies have helped to shape better practices regarding the main issue of vocabulary – which is vocabulary teaching and learning (Laufer, 2003; Nation & Wang, 1999). Vocabulary test tools (e.g. Meara & Milton, 2006; Nation, 1990; Schmitt, Schmitt, & Clapham, 2001) which helped determine L2 learners and natives' vocabulary size have also contributed greatly to the scope of this area of research. They helped in suggesting better judgment of language goals, and in assessing learners' learning needs.

Laufer's (2009) review revealed a number of additional examples of research papers that have influenced SLA research in general. However, it must be noted, of course, that

vocabulary played a partial role as ‘a research lab’ for the main theories during SLA research in general. i.e. the Input Hypothesis (Krashen, 1985), the Output Hypothesis (Swain, 1985), Focus on form (Long 1991) and others. Laufer (2009) gave more attention to this issue in her review, i.e. the impact of vocabulary research on SLA literature. Nevertheless, Schmitt (2008a) states that much of the research has been slow to filter into mainstream pedagogy and argues that there is no best way or method for vocabulary teaching and learning, as practices depend on many factors and ‘‘situations’’ such as the level of learners, the time allocated for learners and the learning goals.

In the following section there is a discussion regarding the main two methods for delivering vocabulary. Several research studies are reviewed, and discussed in relation to the theme of the thesis. However, first, there is a discussion about ‘word knowledge’ and what it entails.

### 2.3.1. Vocabulary knowledge: what does it entail?

Breaking down what it means to know a word has been essential to researchers when proposing the best approach to teaching, applying the most useful assessing tool for measurement, and helping the learning process to be focused. One of the simplest notions about word knowledge is the level of knowing established by Dale (1965): 1) never met it before, (2) heard it but doesn’t know what it means, (3) recognizes it in context as having something to do with, and (4) knows it well (see Beck, McKeown, & Kucan, 2002: 10).

This, however, is a ‘recognition’ level rather than a ‘knowing’ level. Researchers have proposed a precise description of word knowledge (Richards, 1976), a very comprehensive one (Nation, 2001) and a very practical one for educators that describes the learning burden of a word (Nation, 2005). Richards suggested seven aspects of word

knowledge: syntactic behaviour, semantic value, different meanings, underlying form, derivations, associations and limitations on use. Nation (2001) developed Richards' idea further and included more components and a distinction between receptive and productive knowledge as seen below in Table 2.5:

**Table 2.6: What is involved in knowing a word? (Nation, 2001, p. 27)**

| | | |
|---|---|---|
| Form | Spoken | R What does the word sound like?<br>P How is the word pronounced? |
| | Written | R What does the word look like?<br>P How is the word written and spelled? |
| | Word parts | R What parts are recognisable in this word?<br>P What words parts are needed to express meaning? |
| Meaning | Form and meaning | R What meaning does this word form signal?<br>P What word form can be used to express this meaning? |
| | Concepts and reference | R What is included in the concept?<br>P What items can the concept refer to? |
| | Associations | R What others words does this word make us think of?<br>P What other words could we use instead of this one? |
| Use | Grammatical Functions | R In what patterns does the word occur?<br>P In what patterns must we use this word? |
| | Collocations | R What words or types of word occur with this one?<br>P What words or types of words must we use with this one? |
| | Constraints on use | R Where, when and how often would we meet this word?<br>P Where, when and how often can we use this word? |

Word knowledge is not limited to just the 'recognition' of meaning and spelling. Although they are essential for acquisition of words, a learner needs to know much more about lexical items, particularly if they are to be used 'productively' (Nation 2001:27). Nation has, precisely, pinned down how *productive* and *receptive* knowledge may occur in all aspects of a word (*receptive* and *productive* may overlap with *breadth* (how many words known) and *depth* (how well words are known) of knowledge (see the following items).

**Table 2.7 Discovering learning burden (Nation 2005:49)**

| | | |
|---|---|---|
| Meaning | Form and meaning | Is the word a loan word in the L1? |
| | Concept and referents | Is there an L1 word with roughly the same meaning? |
| | Associations | Does the word fit into the same sets as an L1 word of similar meaning? |
| Form | Spoken form | Can the learners repeat the word accurately if they hear it? |
| | Written form | Can the learners write the word correctly if they hear it? |
| | Word parts | Can the learners identify known affixes in the word? |
| Use | Grammatical functions | Does the word fit into predictable grammar patterns? |
| | Collocation | Does the word have the same collocations as an L1 word of similar meaning? |
| | Constraints on use | Does the word have the same restrictions on its use as an L1 word of similar meaning? |

Exploring the word burden, according to Nation (2005) means considering each aspect of what is involved in knowing a word. Table 2.5 above shows the questions that help discover the learning burden of a word.

It has been argued earlier that despite the recent increase of research into second language vocabulary acquisition, the field has tended to lack coherence (Read, 2000; Schmitt, 2008). Two of the main aspects that continue to require further investigation are the issues of how vocabulary is best taught and assessed. For this to be empirically justified, researchers have looked at the aspects of word knowledge to prescribe for which aspect the effective approach to learning and assessment applies. This can be seen from the various works that discuss different perspectives across the word knowledge continuum (starting from the basic recognition levels of the word, and moving to native-like knowledge). Many approaches have been recommended and several types of measures have been proposed (a review of a range of key studies will follow).

Nation and Gu (2007) provided a comprehensive *four-strand approach* which pays balanced attention to learning new information about lexical items, attached with the learning conditions, techniques and activities. The four strands are: (1) meaning-focused input (learning incidentally through reading and listening), (2) meaning focused output (learning by communication through writing and speaking), (3) language-focused learning (form-related aspects), and (4) fluency development (practising what is known already).

Nation (2005) believes that:

"*Learning any particular word should be seen as being a cumulative process where knowledge is built up over a series of varied meetings with the word. At best, teaching can provide only one or two of these meetings. The others involve deliberate study, meeting through meaning-focused input and output, and fluency development activities*."

(ibid: 2)

To sum up, investigating vocabulary knowledge has greatly contributed to the research in L2 vocabulary acquisition, especially to teaching and assessment. Investigating the breadth of knowledge (how many words known) has helped to determine learners' language needs and to decide learning goals (as seen earlier in the chapter). Investigating the breadth of knowledge (how well the words are known) is very useful as it gives a clearer idea about the learners' language capabilities, both receptively and productively (see 2.3.1.2.). This then enables educators to decide effective learning tasks.

However, further investigation of the areas of word knowledge is still essential, especially if different understanding of the issue occurs. For example, *repetition* (or multi-encounter) of words is very effective for better retention of words (Webb, 2007). However, there are only a small number of studies that have explored this, and not all

aspects of word knowledge have been investigated to test the influence of multi-encounter of words. This area also has particular importance in the two main methodologies of vocabulary teaching (incidental and explicit, as discussed in section 2.3.2.). Further research in this area would give a clearer vision about what aspect of word knowledge should be enhanced, and how many encounters (repetition) should be given.

The relationship between the *depth* and the *breadth* of word knowledge still demands further investigations. For example, Gardner (2007) comments on the lexical coverage of texts that "computational corpus based studies may present inaccurate results concerning frequency count." He argues that extra consideration regarding word frequency should be given to some lexical items like *multi words* and *compound nouns*. Words such as *prime minister, break off* and *post man*, would require further analysis, and different perspectives of word knowledge may affect the results that consider *compound nouns* rather than individual items. Martinez (2010) concluded that comprehension levels and coverage of texts varied when considering idioms and multiword items while he was developing the corpus.

Another example is the correlation between the knowledge levels (productive and receptive) of individual learners. Many studies have found that the *receptive* knowledge is greater than *productiveness* of learners (Laufer, 1998; Webb, 2008). Webb (2008) found that "learners who have a larger receptive vocabulary are likely to know more of those words productively'', and of course, "the difference between receptive and productive knowledge increased as the frequency of the words decreased."

In the following section, the relationship between the two knowledge aspects (depth and breadth), with respect to the implication for learners and learning will be discussed.

#### *2.3.1.1. The relationship between size and depth*

The relationship between size (breadth) and depth is not a concept of two distinct points on a line, but rather a dimension within the same circle. Depth and breadth of knowledge overlap, but the first involves quantity of recognition of words and the second involves quality of knowledge. Researching the relationship between the two would be very useful with regards to what aspect of word knowledge should be enhanced (how difficult/ easy it is to acquire), and identify the relationship between the two knowledge dimensions and language competence. For example, as stated in 2.2.1., Qian (2002) believes that scores on depth of vocabulary knowledge can make a unique contribution to the prediction of reading comprehension levels.

In general, as discussed earlier, studies have shown that breadth and depth correlate significantly, and that receptive vocabulary knowledge was always greater (Webb, 2008). The difference between the two dimensions increases with non-common words (ibid). Milton (2009) concludes that many studies show that productive knowledge is 50-80% of the receptive knowledge. The variations of the results are likely, of course, to be affected by the variety of tools used and participants targeted. Administering the vocabulary levels test, VLT, and word associated format WAT (see 2.4.4.) to 44 Korean speakers and 33 Chinese speakers, Qian (1999) found that the score correlated strongly; up to .82 with Koreans and .73 with Chinese. Using the same tools with a larger group (112 Iranian graduates), Akbarian (2010) found a correlation of .74. However, when he divided the students into smaller groups, depending on proficiency, he found that the

higher group reached a correlation of .804 and the lower group reached as low as .464. He suggested that the vocabulary size and depth might account for the same factors, especially as the learners' proficiency increases. This confirms an earlier argument posed by Vermeer (2001) who studied the size and depth of two groups of young monolingual and bilingual Dutch kindergarten learners and concluded that there is no conceptual distinction between size and depth of vocabulary and that they are affected by the same factors for the two groups of subjects (*cf*, ibid). She explained that the greater the number of words known, the deeper the knowledge of the words, and "a child who knows more words also tends to know more about each word" (p. 231, *cf*: ibid).

Richard (2011) compiled a new comprehensive productive test, the (1K-VDT) that measures depth of vocabulary using multiple contextualized examples of sentences (see example below), and linked it to the vocabulary size test (Nation & Beglar, 2007). The participants were 3 L1 natives plus 27 advanced L2 EFL students studying, and in some instances also having previously lived, in an English-speaking environment. Confirming previous studies, the findings revealed that better achievers scored higher in both tests than lower-level students. The correlation was .74. However, the 1K-VDT test would require a threshold of vocabulary size to be able to complete the items. See the example of 1K-VDT in Table 2.8 below:

**Table 2.8 Example of 1KVDT items**

| |
|---|
| (answer: just): |
| It's _____ a small cut. |
| My house is _____ around the corner. |
| It's _____ the right thing to do. |
| We _____ arrived. |
| _____ do it. |
| _____ the two of us. |

1K-VDT (Richard, 2011) is reviewed in Section 2.4.4.

Since researchers state that some productive tests require a threshold of vocabulary breadth knowledge, it is worth investigating how learners would perform in free productive tests that specifically targeted the AWL, and how far results would vary if the AWL were delivered using two approaches.

Webb (2008) has stated that the breadth of knowledge could predict up to two thirds of a person's depth of knowledge. However, the fact that learning vocabulary is an incremental process in nature could imply a variation in dealing with frequency bands (first 1000 words, 2000, 3000 etc.) and the AWL. Although the AWL is included within these bands, it is compiled in a unique list. This hypothesis is supported by the findings of (Chui, 2006)

Chui (2006) targeted the knowledge of the AWL of 186 Hong Kong university students and its link with the productive breadth of knowledge. She measured their vocabulary size by using the Productive Vocabulary Levels Test (Laufer & Nation, 1999) and then sampled 20 words of the AWL and asked students to: 1) construct a meaningful sentence, 2) identify parts of speech, 3) give an assigned derivative, 4) explain the meanings, and 5) select one collocating word out of four options. She found that graduate students know high frequency words, but have a poorer knowledge of the low frequency words. Also, they were able to recognize a reasonable range of the AWL but the quality of this knowledge was deficient (for example they were weak at derivatives and collocations). Chui's (2006) results also mean that achieving 'knowledge' of the AWL does not guarantee a higher 'quality' of knowledge.

Many studies (as reviewed here) have reported similar correlation figures (from and around .80, .74, .73) between the two knowledge concepts even when some had different variables. However, Chui's (2006) study showed that the breadth of knowledge does not provide enough assurance of the level of quality of word knowledge. This would require more studies to investigate specific aspects of knowledge with respect to the vocabulary size. Also, Chui's (2006) study may imply the uniqueness of the AWL or ESP words, compared to larger word frequency bands (e.g. 1000 words, 2000, 5000 etc.). i.e. the studies reviewed earlier (e.g. Milton, 2009; Webb, 2008) state that the general vocabulary size of a learner gives some indication about his/her general vocabulary productive level. This is not necessarily true with the AWL, since it is composed of words from across this frequency range but may contain a large number of items that are outside of the range of knowledge of many learners.

To conclude, investigating depth and breadth of knowledge is crucially important for the teaching and assessment of vocabulary. Research has shown the correlation between the two, and has been successful, to a large extent, in making the link between the two dimensions: with each other, with learners' abilities, and partly, with different frequency bands. However, research is still lacking with regard to the link between specific aspects of knowledge and vocabulary size, and which method of delivering vocabulary is more effective for one aspect or another.

Qian & Schedl (2004: 30) comment:

> '*The importance of various factors in these dimensions will vary according to the specific purpose of language use. For instance, a receptive process may involve a different set of factors than those which may be involved in a productive process*.'

Investigating the depth and breadth with respect to the learning outcomes (resulting primarily from approaches to teaching) would better help decide the learning goals. In the following section, a review of the main two approaches to vocabulary teaching (incidental vs. explicit) is outlined.

### 2.3.2. Approaches to vocabulary learning.

Fortunately, second language learning research is no longer arguing about the priority of teaching vocabulary, but rather discussing the best way of teaching and learning vocabulary.

Although this issue has been frequently investigated, it still requires further analysis in order to confidently recommend better practices for classroom situations. Nonetheless, as has been previously discussed, on-going vocabulary research has been able to set a road map regarding many issues of vocabulary learning that could allow better decisions for classroom practitioners. Concerning teaching approaches for example, research into word frequency, aspects of word knowledge, word assessment, vocabulary size and level of students and others factors would allow for more accurate and justified learning goals to be set, and, accordingly, enable practitioners to recommend what approach is better applied in certain learning situations.

The two main approaches to vocabulary learning in literature are ‘incidental’ and ‘explicit’ teaching of vocabulary. Incidental learning means that the acquisition of words is a by-product of focusing on a spoken and/or a written message, and explicit teaching is when the focus is on lexical items. Sometimes different names are given, but they describe the same concepts; for example, deliberate learning, rich instruction, direct and indirect teaching, contextualized and decontextualized learning. Laufer (2005a, 2006)

proposed *focus on form* in her work involving vocabulary as a response to focus on message agenda of the communicative methodologies. There are several techniques for vocabulary teaching such as; learning by flash cards, dictionary use, online and offline computer application; but they should fall into the main approaches to vocabulary teaching (see Alothman, 2009). Oxford & Crookall's (1990) description of approaches to vocabulary learning might be more accurate and comprehensive. They proposed four categories of the most common techniques of vocabulary instruction; decontextualising (examples: word lists, flashcards, and dictionary use); semicontextualising (word grouping, association, keyword); fully contextualising (reading, listening, speaking, and writing); adaptable: (structured reviewing).

Based on a personal meeting with Norbert Schmitt and on the work of Suhad and Sonbul (2010), it may be better to add *enhanced* vocabulary learning (beside incidental and explicit) to describe explicit teaching of words contained in contextual sentences. At any rate, it could be argued that classifying the approaches into three categories is safer and describes more clearly the actual practices of teaching. To illustrate: teaching words contained in smaller sentences or even written underlined or in bold in a longer text is not incidental, and at the same time could not be described as explicit. *Incidental* learning covers any learning where no focus on lexical items is involved, and *explicit* could mean learning words individually. That is why *enhanced* teaching or *semicontextualising* is thought to be nothing like incidental or explicit (in many occasions in this review, explicit is interchanged with enhanced anyway, see 2.3.2.1.).

Each of these approaches has been discussed thoroughly in the literature and many recommendations and concepts have evolved. The two (or three, based on our argument)

approaches have advantages and disadvantages depending on classroom situations. Alothman (2009) mentioned some of the arguments regarding the two approaches. Here is a summary with some updates:

- Incidental learning matches with the way we acquire our first language; and that is by experience and input rather than direct instruction (Krashen, 1989)
- Context allows better knowledge of words with multiple meaning (Nagy, 1997).
- Incidental learning can address words which cannot be explicitly taught for time reasons, and it can occur while improving other skill areas (Schmitt, 2010b) (among skill areas, reading has been identified in the literature as being the key channel responsible for learners' lexicon increase, both size and depth, as seen above in many studies in sections 2.2.1. and 2.2.2.).
- Incidental learning is seen as a slow means for quantitative word knowledge, compared with explicit instruction (Laufer, 2001; Webb, 2008; Schmitt, 2010b)., and,
- Passive knowledge of vocabulary is weak in instances of incidental learning (Webb, 2007, 2008) (many studies investigated learning vocabulary through classroom talk, living abroad, from watching TV).
- Incidental learning provides more opportunities for words to be met in a variety of contexts (Nation, 2001; Nation & Gu, 2007; Schmitt, 2000).
- Incidental learning provides recycling for words that have been taught explicitly (Nation, 1990; Schmitt, 2010b),
- Incidental learning enriches and strengthens the skill of guessing the word from contexts (Nation & Gu, 2007) and teaches kinds of contextual word knowledge that cannot be easily taught explicitly (Schmitt, 2010b).

- Repeated exposure would guarantee more successful retention (Nation, 2001; Schmitt, 2008a; Webb, 2007).
- Some studies have found that incidental learning has impacted on breadth of knowledge and showed high pickup rates of words encountered in an authentic and systematic reading programme (Horst, 2005; Pigada & Schmitt, 2006).
- Guessing the meaning of a word does not necessarily mean that it has been acquired (incidentally) (Laufer, 2005; Sokmen, 1997).
- Incidental learning does not support the noticing hypothesis (Nation, 2001).
- Explicit teaching involves deeper engagement and consciousness which leads to better retention, robust and faster learning (Schmitt, 2010b) This also includes, of course, the *enhanced* learning of words contained in context sentences.
- Explicit teaching is required for form-related aspects of word knowledge such as affixation, pronunciation and stress, word parts and morphology.
- Matsuoka & Hirsch (2010) investigated how likely an ELT course book is to provide vocabulary learning to learners. They analysed *New Headway Student's Book Upper-Intermediate* and found that it provides opportunities to deepen the learning of the second 1000 words and would provide a context for the AWL words. Their data showed that the text provides fewer opportunities that enhance only a limited part of the AWL, and if the exist, they are not repeated enough. Matsuoka and Hirsch (2010) recommended applying extensive reading programmes plus direct teaching of vocabulary as a supplement.

- People are able to learn between 30 and 100 new words per hour from bilingual word pairs (Nation, 1982 *cf* Schmitt and Schmitt 1995). This shows how fast and effective the explicit method is in targeting individual words.

Scholars recommend a combination of the two approaches in any lexical or general language programme to guarantee enhancing more aspects of knowledge (Nation, 2007a; Schmitt, 2008a). Nation and Gu (2007) proposed the four strands for vocabulary knowledge and suggested the activities that support it (as seen in section 2.3.1). Their activities included learning through incidental input tasks (reading and listening), learning by communication (writing and speaking), explicit instruction of form-related aspects, and rehearsal to develop what have already been learned.

Schmitt (2010b) argues that there is no 'right' or 'best' way to teach vocabulary, but the best practice in 'any situation' will surely depend on many factors. The proficiency level of students, the learning goals and time allocated for learning are some of these factors. Advanced learners are expected to be more capable of learning vocabulary incidentally than beginners, and they can easily guess words within authentic contexts. From personal experience, this, unfortunately, may have resulted in the limiting of the use of explicit teaching of vocabulary to advanced students. Besides, textbooks may not provide enough opportunities for the repeated encounter of vocabulary as reported by Matsuoka & Hirsch (2010), nor support the issue of frequency (Milton, 2009). Therefore, there is some uncertainty regarding the effective approach to teaching vocabulary to advanced EFL learners in presessional courses, and what is best applied in this learning 'situation'.

In the following section, a number of recent studies that have explored the issue of applying incidental and explicit teaching of vocabulary are reviewed and discussed.

*2.3.2.1. Studies on incidental and explicit vocabulary learning.*

Many recent studies have investigated vocabulary acquisition of L2 adult learners. These studies involved unconscious incidental learning, explicit teaching, targeted vocabulary contained within a reading passage, and presenting vocabulary prior to receptive or productive language tasks. In general, explicit teaching was found, as expected, to impact on greater pick-up rate of words. However, some studies revealed the many implications of direct and indirect teaching of words. These are discussed below.

One of the papers that compared incidental and explicit learning and studied their impact on vocabulary knowledge was carried out by Sonbul and Schmitt (2009). They compared vocabulary gains resulting from 'reading only' and 'reading that is aided by teaching some of the technical words'. They assessed three levels of word knowledge (form recall, meaning recall, and meaning recognition) and found that reading plus vocabulary teaching is more effective than incidental learning from reading alone. From this we clearly learn that, with L2 learners, it is crucially important to dedicate some of their EFL learning hours to vocabulary teaching and not depend completely on exposure to vocabulary. However, could this be generalized to a broader situation such as teaching the general AWL list to advanced students? This is especially true if we knew, as discussed earlier, that the current practice of many practitioners of EAP and EAS language programmes does not fully support the explicit teaching of academic vocabulary. Furthermore, since Sonbul and Schmitt (2009) found that presenting the words to learners affected their learning, there is a need to investigate the impact of teaching the general AWL on overall academic skills of advanced learners, if those words were enhanced within the course.

Joe (2010), Folse (2010) and Webb (2009) targeted advanced EFL students, sampling only smaller lexical items. Joe (2010) studied learners' quality and quantity encounters (how many, and it what way encounter occurs) of 20 lexical items over their 3 month EAP course. She found that the frequency of encounters contributes more to vocabulary learning than contextual richness. She concluded that embedding words in rich, instructive contexts on its own did not contribute to better opportunities for vocabulary learning, as it needed to be coupled with noticing and frequently encountering words over a distributed period to improve vocabulary development. She found evidence that shows how noticing a word, having opportunities for focused practice and encountering words over 4–6 days over a distributed period moved one participant's vocabulary development incrementally, even though it was below her optimum threshold of 18 tokens over 9 or more days.

Folse (2010) studied the *explicit vocabulary focus* (EVF) that occurred in a week of classes (25 hours) for one group of upper intermediate students in an intensive English programme. She did an analysis to see if the number of EVF events was connected more with the course (i.e., grammar, reading, composition, communication skills, or TOEFL), the instructor, or both. She found that reading, which for a long time has been assumed to be the source of most vocabulary focus, may in fact not be the main source. She found that the EVF in a week of intensive instruction was surprisingly low. This suggests that relying on incidental teaching in intensive English programmes might be unwise.

Webb (2009) found that the pre-learning of vocabulary positively affected learners' reading comprehension and writing. He taught 15 words using two types of exercises (receptive and productive) to his two groups of advanced EFL Japanese learners (n = 71).

He found that participants who completed the productive learning task had higher scores on the writing test and on the test of productive vocabulary knowledge, while participants who completed the receptive learning task had higher scores on the comprehension test. The study marks on the importance of approach to vocabulary teaching in improving communicative skills. Webb's study, however, targeted a small number of sample words as mentioned earlier. It would be interesting to extend the study to include wider vocabulary or specific frequent ESP or AWL words and explore the impact of teaching those words on ESP or AWL writing or reading.

Elgort (2011) studied the effect of deliberate learning of 48 *pseudowords* on advanced EFL learners' acquisition. She presented the targeted words on cards without their contexts. She concluded that her decontextualized explicit teaching of the words was not only an efficient and convenient but an effective method of L2 vocabulary acquisition, presenting more arguments against the communicative learning methodologies. She added, of course, that this does not mean that words should be acquired exclusively by this method. Elgort's findings reinforces on the effectiveness of explicit teaching of vocabulary to advanced learners. However, she targeted *pseudowords* in her study, not real English words. Teaching non-real words to students raises some ethical concerns, and the permission is not normally granted by school admirations.

It is clear from these studies that explicit teaching is recommended for advanced learners.

Since Folse (2010) and Joe (2010) have used smaller amounts of words, (and sometimes *pseudowords* were used) and found impressive results from explicit teaching, it begs the question; how would learners' performances (receptively and productively) vary if a wider range of words were targeted, involving two methods of vocabulary teaching?

Furthermore, with respect to Webb's (2009), assessment method – productive and receptive exercises-, it is worth investigating how a wider scope study that explores the effect of learning a more relevant vocabulary on learners' academic performance.

Alemi and Tayebi (2011) studied the role of incidental and intentional vocabulary acquisition in addition to the influence of language learning strategy, and surprisingly found no significant variation between using the two types of learning methodologies. They explored the learning outcomes resulting from three methods of presenting new words; *incidental mode, intentional meaning-based mode* and *intentional etymology-based mode* (learning roots, affixation etc.). In the intentional based-meaning mode, words were taught by writing them in bold within the reading passage and the meaning was presented in the footnote. Etymology-based words were taught by analysing them; their roots, prefixes and the suffixes which appeared at the end of the text along with their meanings. The items chosen for the study were 12 words that had been answered incorrectly in the pre-test by all thirty participating students. These words were then included in two lessons before being tested again. Therefore, students learned the intentional words consciously as they paid attention to them, whereas their focus on the content of the reading passage allowed the incidental learning. Learners were told that they would be tested on the words highlighted in bold, but they were not told that they would be also tested on some of the words <u>not</u> in bold. Surprisingly, the results of the t-test of the learners' scores showed no significant difference between these modes. However, a linear regression showed that among the three modes of teaching, intentional meaning-based was found the most predictive of the performance of the learners on the vocabulary test used. These are the means:

Incidental learning; 2.90 (M=2.90, SD=0.95), the etymology-based intentional learning was 2.63 (M=2.63, SD=0.88), and the meaning-based mean was 3.03 (M=3.03, SD=0.80).

Alemi and Tayebi (2011) suspected that some learning attention (i.e. awareness or noticing of the words) must have been involved in the supposed incidental vocabulary, as had occurred in the study of Huckin & Coady (1999). They tried to link the learning process of the students to learning strategy use. No significant correlation was found. However, regarding the procedure of choosing the targeted words for this study, it could be argued that since only a small number of words was targeted (and they were the words that had been answered incorrectly as mentioned above), there might be a chance that low levels of word knowledge could have occurred and incidental learning could have enhanced knowledge. That is why many studies have used *non-words* for their experiments.

To conclude, here are some general arguments:

- Since many recent studies have found that explicit teaching leads to better retention of words than incidental learning, it is worth investigating how the actual application of the two modes would contribute to learners, especially when multiple measures are involved. Besides, the study of Alemi & Tayebi (2011) may have revealed unexpected results in terms of gains from two different outputs. Thus, investigating the issue further and more widely is recommended.
- Since Nation and Meara (2010) recommended graded readers as a means of broadening vocabulary range (specially mastering the 2000 words) and the fact that the AWL words may rarely appear in such books, many studies have recommended extensive

reading programmes for advanced students to enlarge their vocabulary sizes (Horst, 2005; Pigada & Schmitt, 2006). However, they argued that it needs to be systematic and principled in order to guarantee better results. Horst argues that in order to learn the 231(out of 570) most common academic words, learners must read 28 graded readers at level 5 (level 5 is upper intermediate with 2300 words level range). Regrettably, the contact hours of many schools do not fully make use of such reading emphasis.

- Applying extensive reading and extensive listening was found to be a slow process in terms of broadening vocabulary with advanced EFL students in a UK university (AlHomoud, 2007). AlHomoud (2007) argues that direct teaching should enable students to *notice new* vocabulary, either contextualized or decontextualized; whereas extensive reading and listening should help them encounter this vocabulary frequently. Therefore, investigating the intentional teaching of the AWL to advanced students is greatly recommended, to see how it would impact on the learners.
- According to how the words contained in AWL function, it is expected that an even smaller proportion of the list that has been taught to learners would impact much on their skills. Evidence of this was found by Webb (2010). He investigated the pre-learning of the lowest 10 words that appeared frequently in a variety of TV programmes. He found that these 10 low frequent words represented about 0.70 % to 3.91% according to the TV genre. The study revealed that the coverage of 3000 words ranged from 91.83% to 97.83%. If the 10 words were known, the potential coverage would range from 94.93% to 98.66%. What is interesting is that the knowledge of the 10 lowest frequent words gave better potential results than the

knowledge of 5000 words. The 5000 words showed coverage at the 5,000 word level in each TV programme ranging from 0.35% to 1.96%, and was less than 0.70% for five of the eight shows. What contribution, then, would learning the AWL give to the learners' performance?

- Webb (2007) and (Chen and Truscott, 2010) argue that repetition of exposure was found to impact positively on word retention. However, they tested seven aspects of word knowledge (meaning, knowledge of orthographic form, parts of speech, and associations productively and receptively) and found that not all aspects of word knowledge (sometimes meaning) were fully acquired. Chen and Truscott (2010) learned that not presenting the L1 equivalents of words causes learning difficulty. Thus, it would be interesting to trial two methods of teaching the AWL to advanced learners of an intensive course, and assess the impact on aspects of word knowledge.
- Brown (2011) believes that not only do teachers focus too little on vocabulary and enhance different aspects of word knowledge, but that textbooks also do the same. He reviewed nine general textbooks across three levels of proficiency and found that a single aspect – a word's form and its meaning – receives most of the attention in these books. This is an additional reason why a study involving the two approaches would be interesting.
- Research has confirmed that multiple encounters of words increases their retention by learners (for example Laufer & Rozovski-Roitblat (2011). Thus, it is worth investigating how many of the AWL items learners could acquire from the potential encounter within the academic exposure in EAP or intensive courses (i.e. how many items of the AWL are learned in such advanced English courses). Laufer and

Rozovski-Roitblat (2011) even found that some exercises with fewer encounters (2-3) have more of an effect on passive retention than more encounters of other exercises that focus on form (6-7). This finding provides teachers as well as material designers, with useful insights into how to develop more effective exercises.

- Zimmerman (1997a) reported in his research that learners who were exposed to both interactive vocabulary instruction and self-selected reading, ranked activities as more useful to the word learning process than self-selected reading alone.

Finally, it was argued that incidental learning of words within authentic contexts allows better knowledge of words with multiple meaning (Nagy, 1997). Incidental or implicit learning involving substantial communicative input through authentic language seemed to be favoured by educators as the best means to develop proficiency (Folse, 2011). However, it can be inferred clearly from the studies reviewed above in this section that explicit teaching of words is more effective regarding vocabulary knowledge than incidental learning among advanced EFL learners. Therefore, it is interesting to apply actual teaching of vocabulary to advanced learners using these two approaches. However, attempting to conduct such an experiment which involves two groups of similar participants who receive different input would be a brave and challenging move as it might be open to criticism regarding learning and researching ethics. Nonetheless, research into this is still interesting as it would provide practitioners with authentic recommendations about best practices regarding teaching methodologies.

### *2.3.2.2. Teaching advanced learners in a limited time (explicit or incidental learning).*

Although, as has been remarked by (Schmitt, 2008a), there is no clear form of pedagogy in the field of vocabulary teaching, researchers have drawn a road map of it for practitioners and educators. Principles inspired by many studies have facilitated the planning of vocabulary programmes (Folse, 2011; Nation & Meara, 2010; Nation, 2001, 2005; Schmitt, 2008a). Educators need to consider: learners' needs, the words to be targeted, aspects of word knowledge and what to focus on. Two other factors should also be verified as they matter very much: the level and ability of students, and the time assigned to learning. Such factors have, presumably, a significant impact on learning techniques and tasks; the approaches to teaching should be prescribed accordingly. To illustrate, concerning level of proficiency issue, there might be some vagueness concerning the effective methods to vocabulary teaching to advanced EFL learners. Although there is paucity of research concerning the link between methods of teaching words and the level of proficiency or vocabulary size, some efforts have been made to identify any possible associations between them and confirmed that the more advanced the learners are, the more likely they use strategies to learn words (Lawson & Hogben, 1996), learn word meanings from contexts (Carter, 1998; Shefelbine, 1990) and require fewer encounter of targeted words before full mastery (Zahar, Cobb, & Spada, 2001). These results are expected, and it is reasonable to believe that the more advanced the learner, the more likely they are to benefit from learning words in context (i.e. incidental learning). However, this should not give a reason to focus more too often on incidental learning as a source for vocabulary growth among advanced learners, considering the advantages of explicit teaching of vocabulary discussed in section 2.3.2. In fact, in short

or limited EFL or ESP courses, incidental learning of vocabulary could not be the best means for vocabulary development among advanced learners.

Incidental learning of vocabulary seemed to be favoured by educators, as it resembles the natural process for vocabulary development for native speakers (Folse, 2011). This perception could be especially true among many teachers under the influence of communicative language teaching approaches. Furthermore, with advanced learners, expectations concerning the impact of incidental vocabulary learning could be higher, considering the earlier discussions of the results of the three studies (Lawson & Hogben, 1996; Shefelbine, 1990; Zahar et al., 2001) and the argument by Alothman (2011) discussed in section 1.3. However, as scholars remark, what teachers consider useful strategies may sometime be based on assumptions (Carter, 1998; Folse, 2011), rather than theoretical or research grounds. Therefore, there is a need to investigate teaching vocabulary to advanced learners and explore which approach is the most efficient to them considering their level of proficiency and the time allocated for their EFL study.

To conclude, as has been argued earlier, the best practice regarding approaches is to combine, both incidental and explicit learning, where opportunities to enhance different aspects of word knowledge are given to learners. Besides, the learning process will benefit from advantages of the two. i.e. the combination between the two approaches is expected to allow faster vocabulary gains, more successful retention and develop more information about lexical items. This also matches the four-strand approach proposed by Nation (2007a). With the case of advanced learners studying English as a preparation for academic study, a number of factors must be considered in order to achieve a successful vocabulary programme; 1- the vocabulary size of the learners 2- to determine the 'right'

vocabulary to be targeted, 3- the best strategy[s] to be used, 4- that matche[s] with the ability of students, and 5- what is feasible – regarding amount – within the time given.

## 2.4. Vocabulary Testing.

### 2.4.1. The importance of vocabulary testing

The importance of vocabulary testing obviously comes from the importance of vocabulary itself to language teaching. The increased interest in vocabulary research is likely to have implications for vocabulary assessment. As discussed earlier, vocabulary size was found to correlate strongly with language skills. Learners operate better linguistically as their vocabulary size increases. For example, it was mentioned that vocabulary was found to be a good predictor of higher achievement of reading competence (Qian, 2002). This means that vocabulary assessment is essential for research and classroom practice. It provides useful information on how the lexicon of language learners develop, enables researchers 'and practitioners' to know how many words foreign language learners 'know', how fast their target words 'grow' and how these factors are related to other aspects of their linguistic competence (Eyckmans, 2004).

Vocabulary tests, as other language skills assessment tools, can function for different purposes: they help in assigning learners to a suitable learning group (placement test); they can be used to assess what has been learned (achievement test); they can help in detecting the learning gaps in vocabulary knowledge for better classroom planning (diagnostic test); they can give good indications within global measurements (e.g. TOEFL) and allow better estimates of learners' linguistic skills (proficiency testing) (Schmitt, 2000: 164, Eyckmans, 2004: 13)

These assessments, of course, can explore both the vocabulary learners know, and how well individual words are known.

AlSaif (2011) adds that it might be interesting to include vocabulary testing in classroom practice as this would affect the learners' learning behavior and motivate them toward learning vocabulary. He adds that with the increased interest in including vocabulary in the curriculum of EFL courses, there is no point in ignoring vocabulary testing. With the complex nature of vocabulary, the literature actually reinforces including multiple measures while testing (see section 2.4.2. below).

### 2.4.2. The need for multiple measures of vocabulary.

It has been argued on several occasions in this review of the literature that knowing a word is not simply based on a yes/no statement as there are many aspects of word knowledge that should be enhanced for full mastery (see 2.3.1.). Therefore, testing a single aspect of word knowledge might not show a clear picture of the learners' lexical ability. As has been argued before, productive and receptive aspects correlate strongly and some deeper aspects of knowledge give an indication of receptive knowledge (see 2.3.1.2.). Nonetheless, scholars always point out how important it is to apply multiple measures because of the complexity of word knowledge, especially in studies and experiments investigating vocabulary growth (Laufer, Elder, Hill, & Congdon, 2004; Nation, 2007b; Read, 2004; Schmitt, 2010a). Nation (2007b) reviewed a number of studies that applied multiple measures of words knowledge and described how sensitive this is in vocabulary research. For example in Waring & Takaki (2003), three measures were used; knowing that the word occurred in the story, multiple-choice test and a translation test. Not surprisingly, the immediate post-test scores have varied significantly

(15.3 out of 25), (10.6 out of 25) and (4.6 out of 25) respectfully for the three item types. This study implies, according to the researchers, that translating a word is the deepest among the knowledge aspects investigated, and that the recognition of words assumed in multiple choice shows at worst a very good first step towards knowing. Of course, scoring low in a knowledge band does not necessary reflect the depth of knowledge, but at least indicates the variations between levels of knowledge. In other words, the study by Sonbul and Schmitt (2009) discussed earlier in this review found that *form recall* (completion of words aided by initial letters) scored the lowest compared to *translation* and *meaning recognition* (within multiple choice activity). At any rate, the studies highlight the importance and sensitivity of vocabulary knowledge assessment.

Nation (2007b) concludes that the two major pieces of word knowledge are recognition of form, and the connection of this form to a first language meaning. Nation (2007b) has pointed out some other specific advantages from his review of studies that used multiple measures (see ibid: 40). He concluded with yet another advantage for using several test formats of varying difficulty which is "if one of them proved to be too easy or too difficult, then at least the remaining tests will provide usable data" (Nation, 2007b)

Thus, it is highly essential for experimental studies to use multiple measures in any investigations of vocabulary learning. This becomes particularly important in studies investigating vocabulary learning by advanced students studying on ESP and English foundation courses for the following general reasons:

First, as seen clearly from the literature, scholars support the application of multiple measures when testing vocabulary, especially with longitudinal studies (Nation, 2007b; Schmitt 2010).

Second, experimental studies that investigate vocabulary growth of different learners or learner groups exposed to different methods of vocabulary learning are expected to yield sensitive results (ibid). Therefore, using multiple measures should give clearer outcomes that allow for more confident generalizations.

Moreover, using productive vocabulary tests, besides the receptive measure of vocabulary is clearly essential. Scholars see that productive performance implies receptive mastery of words. However, Schmitt (2010: 152) comments that "it is reasonable to assume that…, however, the real danger is making generalizations in the other direction'' i.e. over-estimating learners' productive knowledge relying on their performance of receptive knowledge. Appling different types of productive tests (controlled production/writing tests or free uncontrolled writing tasks) in the studies is recommended in order to overcome any limitations or concerns reported earlier in the literature about each situation (free writing is likely to result in unpredictable results, while controlled tests provide some hints for the test takers, see further discussions below). Further specific arguments will be presented in the review of some of the widely used tests of vocabulary in 2.4.4 below.

### 2.4.3. Reliability and validity of tests: What makes a good vocabulary test?

It is not easy to answer the question of what makes a good vocabulary test, since many issues of vocabulary assessment are still developing, such as frequency levels, the relation between types of knowledge to each other and to proficiency, and maybe other issues in vocabulary research. Researchers may still vary in the vocabulary items they exclude from counts or how they define word knowledge (Milton, 2009). Furthermore, a few standardized tests (see section 2.4.4.) have been released only recently (in the past

two decades) which have made it often impossible to make meaningful comparisons between results of one experiment and another (ibid). Moreover, it would not be very meaningful to compare scores of two types of vocabulary size tests as they would not focus on the same aspects of word knowledge. i.e. scores on a productive vocabulary test cannot be compared against those on a receptive test or vice versa.

For practical reasons, there have not yet been any comprehensive tests that allow testing of every aspect of word knowledge, which include a sufficiently representative sample of words. However, this has not prevented researchers from producing specific tests that were able to give broader estimates of vocabulary size of learners, relying mainly on corpora research and word frequency lists (assembling a representative sample of words from the 1st 1000, 2nd 1000, 3rd 1000 word level etc.). Obviously, good vocabulary tests should be valid and reliable. Milton discussed what this means in his book (2009), as summarized below:

- A test is reliable when it is able to measure something consistently. That includes (test retest method) when a test score of an examinee remains stable after taking two similar tests over a very short period of time; where no significant change would occur in his/her proficiency, and/or the equivalence estimates where different forms of tests compare well and produce equivalent results over a short period of time (Milton, 2009: 17).
- The test is valid when it measures what it is supposed to measure not something else, and when it conforms to the following:

  *Content validity*; whether the test has the necessary and appropriate content to measure what it is supposed to. It should consider the representing words and aspects

> of knowledge. *Construct validity*; whether the test measures the construct or skill it is supposed to (including receptive and productive skills). Measurers of productive knowledge needs to elicit language that is truly representative of learners' productive vocabulary, and requires a method for analyzing this output that fairly and accurately describes vocabulary knowledge. Receptive measures should be easier as they can be controlled by the test creator; however, tasks should be carefully designed to allow students to show the words they know.
>
> *Concurrent validity* is checked when two different tests of the same quality are tried with the same learners. The results of one test should compare well with the results of the second.
>
> *Face validity* checks whether the test is credible to users as a test of what it is supposed to measure. Since some vocabulary tests do not involve explicit vocabulary measurement, some learners may have doubts about the ability for which they are being tested, especially with small scale and simple tests.
>
> (Milton, 2009: 17-20)

Milton's comment concerns the concept of validity in vocabulary testing. O'Sullivan and Weir (2011) and Weir (2005) provided clearer and up to date framework or model for validly of language testing in general. Weir (2005) proposed new frameworks for developing and validating tests of the four communicative language skills. He argues that test developers or users need to address the following questions to offer an acceptable validity of tests. The questions concern these key validity aspects; test taker, context

validity, theory-based validity, scoring validity, consequential validity, and criterion-related validity/reliability:

- How are the physical/physiological, psychological and experiential characteristics of candidates addressed by the test?
- Are the contextual characteristics of the test task and its administration situationally fair to the candidates?
- Are the cognitive processes required to complete the tasks interactionally authentic?
- How far can we depend on the scores on the test?
- What impact does the test have on its various stakeholders? What external evidence is there that the test is doing a good job?

(Weir, 2005: 48)

Building on Weir (2005), O'Sullivan and Weir (2011) presented a practical model that offers test developers a systematic means for generating and interpreting validity evidence during all processes of test development. A graphical representation of their model is shown below in Figure 2-1:

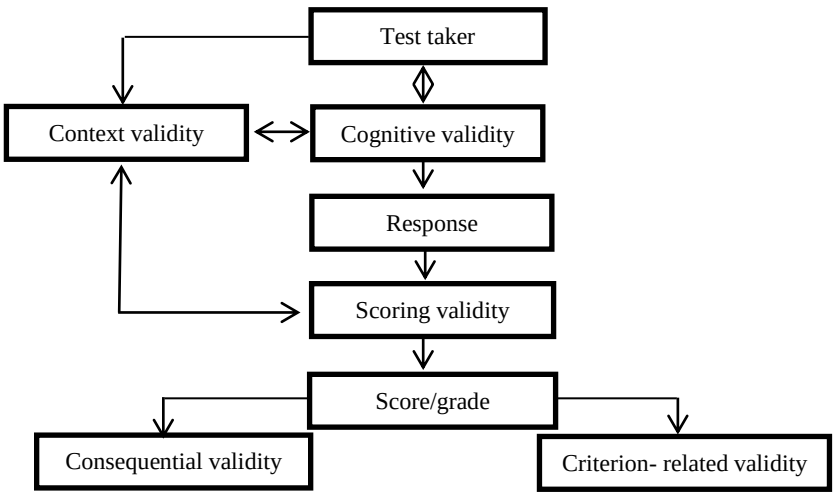



**Figure 2-1: The relation A socio- cognitive framework for test development and validation adopted from (O'Sullivan & Weir, 2011:21)**

With vocabulary, Schmitt (2010a) comments that *criterion* validity is one of the ways to judge a new measurement against a well-established accepted standard test. However, with the complex nature of vocabulary knowledge, some well-known tests may not work as suitable criterion for tests measuring other word knowledge facets (ibid). Schmitt (2010a) discussed a number of global measurement issues in his *Vocabulary Research Manual*. A summary of his discussion and recommendations are pointed out below; some points were found to be relevant to the current study:

- Vocabulary tests can vary along the three *dimensions* of *vocabulary assessment* proposed by Read (2000). The test may either involve assessing *discrete* lexical items or items that are *embedded* within a broader linguistic proficiency, test tasks might involve *selective* items or knowledge or they might assess *comprehensive* output receptively and/or productively, and, lastly, the test items might be *context-dependent* or *context-independent.* Schmitt (2010a) reinforces the fact that testers should be aware of this (as further discussed in the following items).

- When writing *definitions* for *targeted words,* they should be from a more frequent word list to avoid causing confusion to learners. However, this might not always be possible, as some words are hard to define using less sophisticated vocabulary. Interestingly, Schmitt (2010a) found that sometimes dictionaries define some words with even less frequent words. As a second language speaker, it seems clear that some definitions might be even more difficult than the words.

- It is also important that the test items used in the measurement instruments are natural and make sense to the participants. Unnatural test items may occur when contriving

some contexts for lexical items or when developing distractors. Schmitt (2010a) recommends trialing the test with native speakers.

- When selecting target words, testers should consider the nature of some lexical items. For example, some speakers of Romance languages could easily guess some meanings of the words that share similar roots or spelling with the L1. This might not be true for other students from different language backgrounds.

- Delayed post-tests (in addition to the immediate post-tests after a particular study) are recommended in acquisition studies to insure the retention of words.

- Researchers and testers should consider the memory factor when using pre-test/post-test in their work. The inclusion of extra or irrelevant items or tasks, or adding distractors in the pretest would be useful so that the actual target items are less prominent.

- Using different versions of a vocabulary test might be a solution to beat the memory effect; however, it is not an easy job to do too often. Besides, two versions of the same test may not always result in the same score. Therefore, considering larger groups, and/or administering the two versions in the *pre* and *post-test* are recommended to avoid unknown variation due to use of not-fully-equivalent tests.

Schmitt (2010a; 173-179)

Schmitt (ibid) states that vocabulary studies have traditionally focused on receptive vocabulary. Among many reasons which he discussed is that receptive test formats usually offer researchers more control than productive tests, and limit the scope of the enquiry. He comments that ‘free’ output poses difficulties for the tester as the learners’ answers and responses will be uncontrolled and might be unpredictable. However,

choosing to write a word in composition writing could be among the deepest production abilities, unlike the controlled measurements where the test items limit the examinee to particular possible answers (e.g. test items aided by initial letters, or context sentences or definitions). Nonetheless, measuring the frequency of the words produced 'freely' in a composition rather than focusing on specific words can be controlled and assessed. The focus then is on the range of the words within their frequency level bands regardless of the variety or unexpected responses that different writers may produce. Unfortunately, based on the current review of the literature, it seems that not too many researches have addressed this area of study (i.e. free productive ability of vocabulary), most probably due to the challenges remarked above by Schmitt (2010a).

### 2.4.4. Tests that measure vocabulary size.

It was argued earlier that one of the benefits of measuring vocabulary size is to initially check learners' vocabulary capacity for better classroom planning, and observe their vocabulary growth. For this reason, researchers have developed a number of standardised tests that measure vocabulary size. The most well-known test is the *vocabulary level test VLT* developed by Paul Nation (1990). This test was updated by Schmitt, Schmitt, & Clapham (2001) who presented further versions of it, and was validated by Schmitt et al. (2001) and others (see Read, 2007). Schmitt et al. (2001) updated the *Academic Words List* part of the test as a new list of academic vocabulary had been compiled by Coxhead (2000) updating the university word list of Xue & Nation (1984). The VLT has become widely used by teachers as the author made it available in his books (Nation, 2001, 1990) and online[2]. It is a receptive test that requires its takers to match words with their

[2] Vocabulary levels test http://www.er.uqam.ca/nobel/r21270/levels/

synonyms or short definitions (or sometimes L1 translation, see Ishii & Schmitt (2009)). It uses representative words from frequency levels (2,000, 3,000, 5,000, and 10,000) as well as academic vocabulary. The academic vocabulary section could be used apart from the rest of the test items to assess the academic vocabulary only (Norbert Schmitt, personal communication 2009). Table 2.9 shows a sample of the VLT:

**Table 2.9: Sample of the VLT**

| | |
|---|---|
| 1 area | |
| 2 contract | _____ written agreement |
| 3 definition | _____ way of doing something |
| 4 evidence | _____ reason for believing something is or is not true |
| 5 method | |
| 6 role | |

Laufer & Nation (1999) produced a productive test *PVST* that checks vocabulary size by using contextualized sentences and asking test-takers to write the missing words using some lead letters to limit the answers. The authors have trialed the test, presented versions of it and described it as a practical instrument. The *PVST*, however, does not cover items from the academic frequency list by (Coxhead, 2000). Table 2.10 shows a sample of the PVST:

**Table 2.10: Sample of the PVST**

| |
|---|
| 1. The urge to survive is inh……. in all creatures. |
| 2. The ar…….. of his office is 25 square meters. |
| 3. Phil……. examines the meaning of life. |

Nation & Beglar (2007) and Nation & Gu (2007) presented another receptive test- the *Vocabulary Size Test* VST- with a new multiple-choice format, where each target word is presented in a short non-defining context followed by four possible definitions as options

(ibid: 105 and 194). Unlike the VLT and the PVST, it carefully tests representing sample words within their frequency bands (1st 1000, 2nd 1000 to the 14th 1000). The VLT and the PVST test levels in the 2000, 3000, 5000 and 10000 words levels. The test has been reviewed and validated by Beglar (2010). Nation and Gu (2007) have given detailed information about how to use and calculate the scores of *VLT* and *VST*. Basically, in the VST, the score of the 10 items from each 1000 word level is multiplied by 100 to get the total receptive vocabulary size. In the VLT, the score of the 30 items of each section of the levels (2000, 3000, AWL, 5000 and 10.000) is given a percentage against each word level (e.g. score: 22/30= 1466/2000 words, see Scholfield and AlQahtani's formula mentioned below (AlQahtani, 2005), then the total number of words of each section is added to represent a learner's vocabulary size. The sample of VLT is shown in Table 2.9 and the sample of the VST is shown below is Table 2.11:

**Table 2.11: Sample of the VST**

| |
|---|
| 9. **FRACTURE**: They found a **fracture.** |
| a. break |
| b. small piece |
| c. short coat |
| d. rare jewel |

The *VLT* and *VST* provide direct evidence that each word is actually known. However, these tests may have an issue in terms of construct validity as learners need to know the words in the definitions as well as the targeted items (see 2.4.3.). Furthermore, these types of tests give 1 out of 4 chances of guessing the correct answer. For reasons of practicality, these tests cannot be used too often as there are only a few versions

available. It is difficult to sample equal tests with their definitions all the time. Furthermore, as reported earlier, Schmitt (2010a) stated that two versions of the VLT test might not produce statistically equivalent scores with individuals. It was argued before in section 1.3 that the VLT was criticized by AlSaif (2011) for giving overestimations. According to the test scoring scheme, test scores in the 10.000 word level section should not contribute more than the scores of other word bands. See Scholfield and AlQahtani's (AlQahtani, 2005) formula below:

$$\frac{\text{down to the last level sampled}}{\text{Number of items tested at each word level}} = \text{vocabulary size}$$

| | |
|---|---|
| 1- The 2000 word level: | 30/score X 2000 = |
| 2- The 3000 word level: | 30/score X 1000 = |
| 3- The 5000 word level: | 30/score X 2000 = |
| 4- The Academic word level: | 36/score X 836 [a] = |
| 5- The 10000 word level: | 30/score X 4164 = |
| | **The VLT score** = 1+2+3+4+5 |

a. 836 words based on the UWL (Xue & Nation, 1984) which was updated by the formula set by Schmitt and AlHomoud (2007)

**Figure 2-2: Scholfield and AlQahtani's (AlQahtani, 2005) formula approved by Paul Nation**

There is a simpler test that is favoured by many researchers and practitioners and has proved to be reliable and valid (Mochida & Harrington, 2006; Nation, 1990; Pellicer-Sanchez & Schmitt, 2012). This test is the *Check List* or the *Yes/No format*. The form of the test is to sample some words, and learners indicate whether they know them or not. The weakest point of this test is that it depends considerably on learners' behaviour while responding to the test items. However, researchers have included non-words in the test items to control any overestimation by test takers (Anderson and Freebody 1983 *cf* (Read, 2007). To create these non-words or *pseudowords*, the researchers used two principles: 1)

changing one or two letters in a real word and 2) forming unconventional base plus affix combination (Eyckmans, 2004).

One of the most well-known tests in this format is the *Eurocentres Vocabulary Size Test EVST* (Meara & Buxton, 1987; Meara, 1990; and see Read, 2000 for a detailed description and evaluation of the test). It was used by Eurocentre schools as a placement test for learners. It is a yes/no format test that examines the breadth of knowledge of learners of the 10000 frequency word band. 10 real words from each 1000 band are sampled together with another 10 *pseudowords*. It was computerized to make it easier for administering, as the schools run short four-week courses and the test is used frequently. The learners need to click *yes* for the words they think are real and *no* for the words they think are not. If learners score highly at one level, they proceed to the next level. If the score falls to a point within a certain level, the test does not go on to the next level. It estimates that the vocabulary size is somewhere between the previous level and the level where the drop of the score occurred. Although the test works well in terms of its administration, the fact that it stops when the score goes down could raise some concerns about the actual level of vocabulary size of the learners. Learners could actually reach a higher point, so the test should not presume that the learner knows fewer of the infrequent words of higher levels and so does not test them (Milton, 2007).

Masrai (2009) has demonstrated that the EVST resulted in lower scores compared to another paper and pencil *yes/no* test format (the X-lex). He tested 92 graduate students from junior and senior levels in an English Language Department with the following three tests: XK-lex (version A), XK-lex (version B) and EVST. Their mean score in

EVST was lower (4198) than the XK-lex forms A (5370) and B (5186) with seniors, and significantly lower with junior students (EVST: 1680) and (XK-lex: 3109 and 2907).

*The X-lex* (Meara & Milton, 2003) is a computer-based test that checks if learners know the most frequent 5000 words. It presents 100 real words (20 in each band) mixed with 20 *pseudowords*. Scores are calculated by counting the number of *yes* responses to real words and multiplying them by 50, and then the score is affected by any false stating of *yes* to *pseudowords* (a so-called false-alarm). Each false response is penalized by a deduction of 250 points from the overall score. X-lex might underestimate the learners' vocabulary size as it is limited to the 5000 word level band.

*The A-lex* (Milton & Hopkins, 2005) is identical in construction and target size to the X-lex, except that learners listen to the words instead of reading them to indicate their knowledge. The software allows learners to listen more than once before making their decision and this multi-listening does not affect their score.

*The XK-lex* (Meara & Milton, 2006) is another yes/no test format. However, it solves the issue thought to cause the underestimation of learners' vocabulary size proposed earlier in both the EVST and X-lex. Since it comes in pencil and paper format, learners can access all levels regardless of any poor performance at any band level, unlike the EVST. The EVST resulted in lower scores for learners compared with XK-lex as seen in Masrai (2009). Concerning XK-lex target words, it covers up to the 10,000 word frequency band, unlike the X-lex which only tests below the 5000 word level, and, therefore, providing less opportunities for advanced students to score higher. We have seen, according to Masrai (2009), that the mean score of the XK-lex reached 5370 words, and for some of his learners the maximum score was 7800 words. AlSaif (2011) tested students of a

similar level and found that the mean was 3.252 words and the maximum was 4850 words. He surmised that X-lex and A-lex might limit the estimation of the abilities of advanced students.

Like the X-lex, the XK-lex presents 120 words (100 real words and 20 *pseudowords*); however, only 10 words represent each frequency band instead of 20 words in each band in the X-lex. In working out the raw score for participants, each real word chosen as known is scored 100 marks, and the final score deducts 500 marks for each unreal word chosen as known. Masrai (2009) has carefully checked two versions of XK-lex for validity and reliability and found it to be a consistent and valid measure of breadth of vocabulary knowledge. For a sample of XK-lex see the Appendix A and B.

*The Computer Adaptive Test of Size and Strength (CATSS)* Laufer & Goldstein (2004; Laufer et al., 2004; Laufer, personal communication) aims to assess vocabulary size, i.e., knowledge of word meaning, by applying four modalities to test the strength of the aspect of meaning of each word. Here is an example of how each type of knowledge can be elicited for the same item: the word *melt:*

*Type 1: Active recall*

It asks for the word to be supplied with the help of an initial letter to avoid non-target words:

1. general idea used to guide one's actions p_____

*Type 2: passive recall*

It presents the word in an embedded phrase and asks the testee to supply an acceptable answer that shows the meaning (e.g. beliefs, rules, ideas, basic rules, basic ideas):

If you're guided by a *principle*, you follow some _____

*Type 3: Active recognition*

The task is to choose the target word from four options

General idea to guide one's action

a. fund b. percent c. principle d. philosophy

The options are from the same frequency band (in general vocabulary), so learners could know the answer by eliminating the incorrect answers which also represent this level.

*Type 4: Passive recognition*

Principle

a. study of the meaning of life b. money for a special purpose

c. general idea to guide one's actions d. one out of 100

Here the test taker chooses the meaning from the options which may sometimes be expressed in the L1.

The computerized version of CATSS was proposed by Laufer et al., (2004) who looked at the relations between all levels of strength of knowledge in the meaning aspect. It runs on a formula based on a hierarchy of size and strength modalities. For example, if a learner gets an item correct in one strength modality, it will not be necessary to test the same word in the subsequent strength modalities. If the item is not answered, the test keeps the word in memory to be presented in another modality. Batia Laufer (personal communication 2010) published a paper and pencil version of the test. She states that if the pencil and paper of CATSS is applied, it is better not to give all tests immediately after each other as learners might be able to remember the words from previous test type.

*The vocabulary profiler (vocabprofiler)* (Cobb, 2008) is a web tool that checks Lexical Frequency Profile of learners' written production, i.e. looks at the proportion of high frequency general service, academic words, as well as words from the British National

Corpus High Frequency Word List. It was inspired by ***Range***, an offline software and a well-known research tool developed by Laufer and Nation (1995). *Vocabprofiler* has many features and can count, in various ways, the number of *tokens*; all words in texts regardless of how many times they appear, number of *families*; headword[s] and its inflected forms are members of one family word, and *types*; different forms of a word that appear in the text taken together (play and player are two different types) (Milton, 2009). Regarding the AWL, the software can target the words which appear in the text and process them according to their frequency sub-lists within the AWL.

*Word Associates Test (WAT)* (Read, 1993) is a vocabulary depth of knowledge test and it measures three vocabulary elements: synonymy, polysemy, and collocation. Validated by Read (1993) and Qian (2002), it has 40 items (objectives), each of which has two boxes containing four words. The test states that there are four possible answers out of the eight words.

**Table 2.12: Sample of the WAT**

| |
|---|
| • Critical (choose four words that associate with critical) |
| clear dangerous important rough // festival illness time water |

The test does not target the whole range of the academic vocabulary list.

The (1K-VDT) was compiled by Richard (2011) as mentioned above in section 2.3.1.2. It is a productive vocabulary test that measures depth of vocabulary using multiple contextualized examples of sentences. The learner answers by producing the word that fits all statements. Below is an example:

**Table 2.13: Sample of the 1K-VDT**

| |
|---|
| *Fill in each of the blanks with one word.* |
| She held the young boy in her **arms .** |
| Matsuzaka has a good **arm .** |
| As they walked, he offered her his **arm .** |
| The political **arm** of the group met with the media. |
| Both sides agreed to **disarm .** |
| Mom **armed** us with supplies to get the house ready. |

The 1K-VDT test was compared to the receptive VST test (Nation & Beglar, 2007) and showed significant correlation, indicated by the variation between higher level students and lower level students. Table 2.14. below summarizes the findings:

**Table 2.14: Descriptive statistics for two groups on two tests (Richard, 2011: 113)**

| Test | Higher (n=20) | Lower (n=10) |
|---|---|---|
| VST | mean = 88.30<br>St Dev = 11.76<br>max = 107<br>min = 67 | mean = 77.20<br>St Dev = 5.99<br>max = 91<br>min = 72 |
| VDT | mean = 44.65<br>St Dev = 4.39<br>Max = 49<br>Min = 32 | mean = 41.23<br>St Dev = 3.30<br>max = 47<br>min = 37 |

Richard (2011) however, reminds us that the VST is different from the VDT as each measure is meant to test a different aspect of knowledge. Richard (2011) revealed interesting results regarding the link between size and depth as previously discussed in section 2.3.1.2. However, the number of items in his study was quite small, and therefore, care is needed in interpreting his findings. Regarding the *face validity* of the test, there may be a concern this issue and how the test functions. In other words, one sentence of each of the six items of the test could give the rest away. O'Sullivan (2011, personal

communication) believes that if the test was computerized and all the answers of the items (plus distractors) put in a dropdown menu, with the sentences mixed up rather than divided into sets of six each, the test could work well regarding this *validity* issue. The test does not fully support the AWL.

## 2.5. Summary and conclusion.

Vocabulary is essential for language competence and was found to correlate significantly with the linguistic performance of L2 learners, for example, the knowledge of even a small number of words was found to considerably affect comprehension (as seen in Webb, 2010). Therefore, learning academic words or English for specific purpose (ESP) words is beneficial for EFL students. However, advanced students were often thought to be capable of the incidental learning of vocabulary through reading and listening, therefore, very little explicit vocabulary instruction tends offered to them in their language programmes, as discussed in section 2.3.2.2.

Folse (2011) argues that, since native speakers do not learn most of their vocabulary through explicit instruction, common wisdom in ELT pedagogy has favoured a natural approach involving substantial communicative input through authentic language as the best mean to develop proficiency. Particularly, with advanced or intermediate EFL learners, this view toward incidental learning has become more popular among educators, considering the expected positive association between learners' level of proficiency and the ability to learn words from contexts (as discussed previously in section 2.3.2.2). However, this does not necessarily justify neglecting or paying less attention to explicit vocabulary teaching to advanced learners. As reviewed in section 2.3.2.1, a number of studies (Alemi & Tayebi, 2011; Elgort, 2011; Folse, 2010; Joe, 2010; Sonbul & Schmitt,

2009; Web, 2009) have shown the benefits of the direct teaching of vocabulary to advanced learners over indirect or incidental learning. There are a few limitations, however, concerning some of the studies that this paper reviewed regarding the type of experiments being carried out (Alemi & Tayebi, 2011; Sonbul & Schmitt, 2009; Webb, 2010), the words targeted (Elgort, 2011; Folse, 2010; Joe, 2010; Sonbul & Schmitt, 2009), the type of measurement used (Webb, 2009), and ability to generalize the results to any great extent (Alemi & Tayebi, 2011; Sonbul & Schmitt, 2009).

Thus, with respect to the above arguments, it is essential to research further teaching vocabulary to advanced learners, and investigate the learning outcomes resulting from applying incidental and/or explicit approaches to vocabulary learning. It is important to identify whether or not advanced learners can rely on incidental vocabulary learning because of their advanced linguistic abilities, or should they learn vocabulary through direct instruction. Of course, in an attempt to undertake such an investigation, it is important to assess different aspects of word knowledge due to the complexity of vocabulary acquisition that has been discussed extensively in this review (see sections 2.3.1 and 2.4.2). The use of multiple measures of vocabulary knowledge could present, to a great extent, clear results about which method of vocabulary teaching would be more beneficial for advanced learners. Furthermore, it has been identified in the above review that advanced EFL learners are perceived to be capable of learning vocabulary implicitly, and therefore, less direct vocabulary instruction is applied to them. On the other hand, some studies involving learners' feedback and views about learning reviewed earlier (section 2.2.4) report that learners, in fact, state their needs for vocabulary teaching. Unfortunately, many of the studies, as discussed, addressed learners' views about general

language learning including vocabulary. Thus, it is important to investigate advanced EFL learners' beliefs about learning vocabulary, and report their feedback about the factors which might influence their learning. Lastly, the academic word list is considered to be highly essential for learners of English at the academic level as discussed in section 2.2.3. However, with respect to the above arguments about common perceptions towards advanced EFL learners abilities, and the complexity of vocabulary knowledge and need for multiple testing, this review of the literature revealed some uncertainty about the level of advanced learners at the pre-university level regarding the mastery of the AWL (as discussed in sections 1.4 and 1.5.). All of this indicates that there is a real demand for researching the exact level of advanced learners regarding the mastery of the AWL, and investigates whether relying only on incidental learning of vocabulary is actually enough to enhance different aspects of knowledge.

## 2.6. Hypotheses and research questions of the thesis:

The above literature review detailed some theoretical background about the needs of advanced EFL learners with regard to academic vocabulary, approaches for vocabulary learning among advanced adult learners, dimensions of vocabulary knowledge and mastery, and learners' and teachers' beliefs about vocabulary learning at an advanced stage of EFL study. Emerging from this review is a series of research areas that appear to demand further investigation:

a) The learning of academic words among learners at advanced EFL programmes, considering their actual (or general) vocabulary levels, and the complex nature of vocabulary knowledge.

b) The variation in terms of academic vocabulary gains between incidental or indirect learning of the AWL and explicit teaching of the words, considering the complex nature of vocabulary mastery.

c) Learners' and teachers perceptions and beliefs about vocabulary teaching at this level of proficiency.

Thus, with respect to the need and motivation for this research discussed in sections 1.4 and 1.5 above, and the theoretical arguments and concerns raised in section 2.5 and elsewhere within this chapter, this study is designed to explore the three areas mentioned above. To do this a series of research questions and related hypotheses have been devised and are presented in Table 2.15.

**Table 2.15 Summary of the research questions and the related hypotheses:**

| **A) The learning of academic words among learners at advanced EFL programmes, considering their actual (or general) vocabulary levels.** | |
|---|---|
| **RQ 1-** How does the depth and breadth of academic vocabulary (as measured by the AWL) of learners at an advanced stage of their pre-university studies change during their preparatory programme?<br><br>**Sub-question**<br><br>What is the relationship between academic vocabulary, breadth and depth, and general vocabulary size of learners? | **RH 1-** Learners at an advanced stage of an academic preparatory programme will gain significantly ($\alpha$**<.05)** in their receptive knowledge of academic vocabulary.<br><br>**RH 2-** Learners at an advanced stage of an academic preparatory programme will gain significantly ($\alpha$**<.05)** in their productive ability of academic vocabulary.<br><br>**RH 3-** There is a significant correlation between receptive and productive AWL abilities, and general vocabulary size. |

**B) The variations in terms of academic vocabulary gains between incidental or indirect learning of the AWL and explicit teaching of the words.**

| | |
|---|---|
| **RQ 2-** Which approach to vocabulary teaching (incidental/explicit) is likely to lead to the most significant gain?<br><br>**Sub-question**<br>To what extent do different methods of AWL learning affect the different abilities or aspects (recognition, production and general vocabulary size) of lexical knowledge? | **RH 4-** Explicit teaching will lead to more significant gains in the receptive knowledge of the AWL than incidental learning.<br><br>**RH 5-** Explicit teaching will lead to more significant gains in terms of the productive knowledge of the AWL than incidental learning.<br>**RH 6-** There is a significant correlation between the method of teaching the AWL and the different abilities or aspects (recognition, production and general vocabulary size) of lexical knowledge being measured. |

**C) Learners' and teachers perceptions and beliefs about vocabulary teaching at this level of proficiency**

| | |
|---|---|
| **RQ 3-** What are learners' and teachers' perceptions and beliefs about vocabulary teaching at this level of proficiency. | **RH 7-** Learners and teachers will be positively disposed towards the explicit teaching of academic vocabulary in their advanced EFL programmes. |

***<u>First</u>: The learning of the AWL receptively and productively among pre-university learners within their advanced EFL programmes***

Based on the discussions in section 2.2.3 about the variations in the figures of vocabulary thresholds recommended for EFL learners suggested by Nation (2006) and followed up by Laufer and Ravenhorst-Kalovski (2010), the uncertainty about the capabilities of advanced EFL learners regarding the AWL (Chui, 2006) and the concerns regarding differences between receptive and productive knowledge dimensions (Milton, 2009) discussed in section 2.3.1.1, the limitations regarding vocabulary levels sizes discussed in section 1.3 and reported in Schmitt (2008a), and the complex nature of vocabulary knowledge discussed extensively within this chapter, the main research question (RQ1)

addresses the issue of: How many words of the AWL could learners at pre-university level recognize, and able to produce at this advanced level of EFL proficiency?

Regarding RH 1 and 2, expectations are that learners would improve their receptive and productive abilities of the AWL due to their advanced level of language study, which would also support the findings of the studies of incidental learning of words discussed earlier in section 2.3.2.1 (Horst, 2005, 2010; Nagy, 1997; Pigada & Schmitt, 2006; Waring & Takaki, 2003). However, learners are expected to learn more AWL words receptively than productively based on the earlier discussions about the complexity of word knowledge. Many studies reported in section 2.3.1.1 (Akbarian, 2010; Qian & Schedl, 2004; Qian, 2002; Vermeer, 2001; Webb, 2008) confirm the significance of correlations between receptive and productive abilities, and Milton (2009) concludes that many studies show that productive ability tends to be 50-80% that of the receptive knowledge. However, this may relate to frequent general words, and might not be true for less frequently occurring words such as the AWL or ESP words, as Chui's (2006) study implies.

This main research question (RQ1) imposes a further additional research query ARQ which addresses:

> ***ARQ:*** ***The relationship between learners' recognition and production of the AWL, and their general vocabulary size.***

This additional sub-research question reflects the complexity of word knowledge discussed extensively in this thesis, the interaction between the receptive and the productive vocabulary knowledge discussed in section 2.3.1.1, and considers the different

measures that assess the AWL which the above main research question addresses (i.e. recognition, and production of the AWL in free and controlled writing tasks).

Regarding the vocabulary size, since the AWL fits in a broad range between the 2000 level and the 10,000 level (Schmitt et al., 2001), and is contained within the 10,000 word levels measures (e.g. VLT and VST), the studies that explore overall scores of vocabulary size (e.g. AlHomoud, 2007; AlNujaidi, 2004; AlQahtani, 2005) give broad estimations of vocabulary size including the AWL frequency to work out the overall score (see Figure 2.2 for illustration). There could be some vagueness about specific AWL levels as the overall score could not give accurate indications about the knowledge level regarding this band. Therefore, with respect to the multiple measures used to assess the AWL addressed in the main RQs mentioned above, the sub-research question also explores the relationship between learners' abilities (receptively and productively) in using the AWL and their overall vocabulary size.

Regarding RH 3 concerning the additional sub-research question, expectations are that the scores of the two AWL tests will vary significantly, considering the variations between the two aspects of knowledge discussed in the studies in section 2.3.1.1. Recognition of the AWL is expected to be higher than the ability to produce it, based on the findings of many studies discussed in section 2.3.1.1 (e.g. Webb, 2008; Milton, 2009). Concerning the general vocabulary size and its relation to the two different AWL abilities, it is expected that the two different AWL test scores correlate significantly with general vocabulary size, since the academic list is contained within the 10,000 most frequent common English words, as discussed above. With respect to the two different AWL aspects of knowledge, it is expected that the two dimensions would interact

differently with the general vocabulary size, considering the variations between the knowledge mainstreams.

***Second:*** ***Learning the AWL by learners exposed to two different methods of teaching (incidental and explicit teaching/learning)***

RQ2 reflects on the discussions and arguments raised in sections 2.2.4, 2.3.2.1 and 2.3.2.2 regarding approaches to vocabulary teaching (incidentally or explicitly) to advanced EFL learners (AlHomoud, 2007; Folse, 2010; Horst, 2005, 2010; Joe, 2010; Nagy, 1997; Pigada & Schmitt, 2006; Sonbul & Schmitt, 2009; Waring & Takaki, 2003; Webb, 2009), considering the complex nature of vocabulary knowledge that demands multiple testing.

Concerning the first RH (RH4) related to this question, expectations are that learners exposed to direct enhancement of the AWL would gain more words receptively and productively than learners relying on incidental vocabulary learning based on the discussions in 2.3.2.1 (Akbarian, 2010; Folse, 2010; Joe, 2010; Sonbul & Schmitt, 2009). There might be some uncertainty about the degree of variation between the results from both learning approaches with respect to the different word knowledge dimensions, but would be generally in favour of the explicit teaching learner group.

This main research question (RQ2) imposes a further additional research query ARQ which addresses;

> ***ARQ:*** ***The effect of the method of AWL teaching (incidentally or explicitly) on the different mastery levels of the AWL.***

This additional question explores how far different methods of AWL learning affect the different abilities or aspects of lexical knowledge by learners experiencing those two different methods of learning academic vocabulary (i.e. incidental or explicit teaching/learning).

Regarding RH 6 concerning this additional sub-research question (ARQ2), expectations are that the different abilities and knowledge regarding the AWL would vary according to the methods of teaching the academic list (AlHomoud, 2007; Folse, 2010; Horst, 2005, 2010; Joe, 2010; Nagy, 1997; Pigada & Schmitt, 2006; Sonbul & Schmitt, 2009; Waring & Takaki, 2003; Webb, 2009). I.e. based on the discussions in section 2.3.2.1, learners exposed to direct enhancement of the AWL would score relatively higher in all different tests than learners learning the AWL incidentally.

***Third: Learners' and Teacher' feedback on learning the academic word list***

Research question 3 reflects on the discussions and concerns raised in section 2.2.4 regarding the studies using the *Beliefs about Language Learning (BALLI)* that addressed general language skills including vocabulary, and the comment by Alothman (2011) discussed in 1.4 regarding perceptions of EFL teachers towards teaching vocabulary to advanced learners, as well as the current argument raised in section 2.3.2.2 regarding the common perceptions among educators concerning the method of vocabulary learning best applied to advanced learners (Carter, 1998; Folse, 2011; Lawson & Hogben, 1996; Shefelbine, 1990; Zahar et al., 2001). It explores learners' beliefs and perceived needs regarding the AWL learning on advanced EFL intensive courses. It explores how learners and their teachers feel about the way vocabulary is taught on the courses they are involved in.

Concerning the RH regarding this question, it is expected that learners state their needs and confirm their interest in vocabulary teaching, regardless of the fact that learners are considered advanced and therefore viewed to have less interest into vocabulary learning, as argued by Alothman (2011) and commented by Folse (2011). These expectations are supported by the results in the limited studies discussed in section 2.2.4 that involved questionnaires of general *Beliefs About Language Learning Inventory (BALLI).*

Three independent research studies were designed based upon the three research questions mentioned above.

The main research questions of each study of this thesis and the additional queries are repeated briefly in section 3.3 with respect to each study described.

# 3. Chapter Three

# Methodology

## 3.1. Overview of Chapter Three

This chapter describes the methodology undertaken in the three studies of this thesis: 1- investigating the learning of the academic word list on typical presessional courses in the UK, 2- investigating the learning of the academic word list by learners experiencing two different methods of teaching on an advanced intensive English programme, and 3- learners' and teachers' feedback on learning the academic word list. The chapter first presents an overview of the purposes of each of the three studies of the thesis and what each study is expected to contribute. It then briefly re-presents the main research questions of each study of the thesis, then discuses the additional research queries that emerge. It continues by describing the general design and methodology used in these studies, and how the studies differ in terms of their aims and research questions. It then discusses the main instruments used in the studies, justifying why they were used or modified, according to their aims and settings. A description of the participants and the background of the studies is presented, followed by a description of the general procedure of the analysis of the data in the three studies. The chapter concludes with a summary of the organization of the remaining chapters.

Detailed descriptions of the methodology of each study are in Chapters 4, 5, and 6.

## 3.2. Description of aims and purposes of each of the three studies of the thesis:

Vocabulary has long been considered as the most essential and critical linguistic component. Vocabulary size has been found to have a direct link with language skills; reading, writing, speaking and listening (as discussed in section in 2.2.1.). For example,

research has found that vocabulary knowledge contributes positively to comprehension (e.g. Laufer, 1997, Nation, 2006). More interestingly, Webb (2010) has found that the knowledge of even a smaller proportion of words (the least frequent 10 word items in a targeted genre) was found to considerably affect comprehension. Thus, learning the academic words that appear frequently in academic contexts will undoubtedly be useful. Unfortunately, as discussed in section 1.4., many advanced EFL students at the academic level fail to reach the 2000 word level, which consequently means they have poor knowledge of academic words. The case of Saudi graduates might be seen as even more challenging. Studies concerning the Saudi context reviewed in Sections 1.4 and 1.5 report lower vocabulary sizes among Saudi EFL learners at the pre-university level. Their vocabulary size varied from between 1000 and 3000 words, depending on the type of the test used, and level and major of students. However, even with the optimistic figures reported in these studies, the current research had some concerns about the ability to generalize some of the figures (see the discussion in section 1.5.). At any rate, the studies report poor knowledge of the academic words at university level. Thus, the inclusion of teaching of the AWL in the academic curriculum would be a wise decision.

Regarding the method of teaching vocabulary, advanced learners have long been perceived to be more capable of learning language input implicitly and, therefore, less direct vocabulary instruction is required (see discussion in section 2.5.). Folse (2011) comments that since native speakers do not learn most of their vocabulary through explicit instruction, common wisdom in ELT pedagogy has favoured a natural approach involving substantial communicative input through authentic language as the best means to develop proficiency. However, several studies have been reviewed in this thesis that

show that direct teaching of vocabulary to advanced students is hugely beneficial, regardless of the limitations that have been noted in sections 2.3.2, 2.3.2.1 and 2.4.2.

Thus, with respect to the issues raised above, the three studies were designed to explore the following issues.

*Study One:*

With the impact of the AWL on language skills development that has been discussed on several occasions in this paper, and the fact that advanced learners are thought to be more capable of grasping language input through exposure, this study investigates how many of the academic words advanced students at typical UK presessional English programme know, and how well they know them. Additionally, the study investigates how many of the academic words they learn during the programme, and how well. Also, since the current thesis uses multiple measures to test words, the question of how learners vary in terms of their abilities and knowledge of the academic words is considered.

*Study Two:*

As discussed several time in this thesis, and argued in Alothman (2011), educators may seem to favour implicit teaching of language to advanced learners due to their abilities compared to lower level students. Therefore, this study explores how learners involved in two methods of learning the AWL (implicit learning vs explicit teaching) would vary in terms of their gain of words. Multiple measures to test vocabulary are used in this study to investigate how students of the two focus groups vary in terms of their knowledge of different aspects of a word.

*Study Three*:

Regarding what has been mentioned earlier about the importance of academic vocabulary for advanced learners and how the AWL is best learned, the third study reports the views of both advanced learners and teachers about the AWL, in terms of learning needs and method of learning.

### 3.3. The main and additional research questions of the three studies

The main research questions RQs and hypotheses RH were discussed in Chapter two. As mentioned, the main research questions of this paper imposed further additional research queries that serve the aims of each study presented in this thesis. A brief repetition of all research questions with respect to each relevant study is mentioned below, and then this is followed by a description of the further additional queries for each study

*Study One:*

| |
|---|
| Investigating the Learning of the Academic Word List on typical presessional course in the UK. |

Research questions of Study One:

RQ 1- How do the depth and breadth (recognition/production) of academic vocabulary (as measured by the AWL) of learners at an advanced stage of their pre-university studies change during their preparatory programme?

*Sub-question*

- What is the relationship between academic vocabulary, breadth and depth, and general vocabulary size of learners?

*Study Two:*

| |
|---|
| Investigating the Learning of the Academic Word List by learners exposed to two different methods of teaching in an Advanced Intensive English Programme. |

Research questions of Study Two:

RQ 2- Which approach to vocabulary teaching (incidental/explicit) is likely to lead to the most significant gain?

*Sub-question*

- To what extent do different methods of AWL learning affect the different abilities or aspects (recognition, production and general vocabulary size) of lexical knowledge?

*Study Three:*

| |
|---|
| Learners and Teacher' feedback on learning the academic word list. |

There are two research questions for this study:

RQ 3- What are learners' and teachers perceptions and beliefs about vocabulary teaching at this level of proficiency?

The main research questions and the additional queries obviously involve determining the vocabulary levels of learners in all different tests at the beginning of their programmes, their achievements at the end, and the degree of development during the targeted duration.

## 3.4. General design of the three studies of this thesis:

Three research designs have been set up to serve the purposes of each study of this thesis. *The first* and *second studies* are similar to each other regarding the general design, assessment timing, and procedures. However, depending on their research objectives, the study designs vary in terms of grouping and treatment, data analysis, and some of the

assessment tools that have been adjusted according the purpose of the study, the participants, and the context.

The third study (*Learners' and teachers' feedback on learning the AWL*) shares the same participants and settings of *study one* and *study two*. i.e. at the end of each of the first two studies, learners and teachers have recorded their feedback concerning many issues of learning behaviour, learning outcomes, abilities and learning needs of the AWL. This was treated as an independent study as it covered more issues and research questions regarding learning the AWL, apart from the feedback. A further discussion about this is in Chapter Six (see 6.1)

The general research designs and procedures that were carried out in the three studies are outlined below. More detailed information about procedures, participants and research tools of each study are mentioned in *Chapter* 4, *Chapter* 5 and *Chapter* 6.

*Study one*: *Investigating the Learning of the Academic Word List on typical presessional course in the UK*.

A group of learners of presessional courses (see Section 4.2.1) were assessed with four measures (see Section 3.4.1.) to test their academic vocabulary abilities. They received the four measures twice - as pre- and posttests - over their 10-week presessional programme. The aim as mentioned earlier was to evaluate the advanced EFL learners' abilities in using the AWL on typical presessional programmes, and to measure their normal (typical) improvement during the programme, using multiple measures.

*Study two*: *Investigating the Learning of the Academic Word List by learners exposed to two different methods of teaching on an Advanced Intensive English Programme*.

This study, as mentioned earlier, is similar to *the first study.* However, this study aims to assess the academic vocabulary abilities of advanced EFL learners receiving two different exposures to the AWL during their EAS/ESP language programme. In this study, similar learners of two learning groups were selected to take part in the research project. Both groups received normal academic exposure in their EAS programme, however, one of the groups (the treatment group or the direct group, DG henceforth) received an extra outside classroom activity that enhanced the AWL (there might be some concerns about some of the ethical issues of grouping and treatment; see the discussion in section 5.2.2.).

Similar to *study one,* the participants of both groups took the four vocabulary measures at the beginning of their term-one (2nd week) and once more towards the end of their course ($10^{th}$ week).

*Study three*: *Learners' and Teachers' feedback on learning the Academic Word List.*

This study seeks perceptions about learning the academic words by advanced learners as well as their teachers. It also intends to tackle some of the research gaps and limitations that have been discussed earlier about learners' needs, methods of learning vocabulary and teachers' perceptions (see as an example the discussions in sections 2.2.4. and 2.3.2.2.).

The participant students of the first two studies as well as a number of their teachers were surveyed for the study. The students answered a 5-point scale questionnaire, and the teachers answered an open-ended questionnaire (more information in Chapter 6). Both learners and teachers were asked to declare some personal background information.

Table 3.1. below summarizes the general procedures of the three studies. A column that states briefly the method of data analysis is added to the table, and further details are in section 3.4.3. The research instruments are presented in the following section.

**Table 3.1 Summary of the general procedure of the *first study* and the *second study*:**

| Research question | Instrument [a] | Participants | analysis |
|---|---|---|---|
| Receptive knowledge of the AWL among advanced EFL on presessional English course | The VLT version **A** and **B** | All participants | t-test [b] |
| 1st AWL Productive knowledge of the among advanced EFL on presessional English course | Adopted version of ***CATSS***-*same test twice* | All participants | t-test |
| 2nd AWL Productive knowledge, whether knowledge of the AWL by some learners affects their lexical choice in free academic writing | Writing Task **A** & **B** and vocabprofiler | Representative | Vocab-profiler |
| Sub-RQ: Overall general vocabulary size of learners | Version **A** and **B** of XKlex | All participants | t-test |
| Sub-RQ: Relationship between the AWL measures to each other and to the vocabulary size | All measures | All participants & Representative | Manova qualita-tive |

[a:] *instruments described in section 3.4.1*
[b:] *or equivalent depending on normality of distribution. Instruments*

**Table 3.2 Summary of the general procedure of study three: Learners and teacher' feedback about learning the AWL:**

| Research question | Instrument | Participants | analysis |
|---|---|---|---|
| Learners: What are learners' beliefs as well as needs regarding the AWL in intensive courses? | 5-point scale questionnaire | Students | Correlations |
| Teachers: What is learners' and teachers' feedback about the way vocabulary is taught on the courses they are involved in? | 5-point-scale questionnaire *plus* Open-ended Qs | Students and teachers | Qualitative and Correlations |

### 3.4.1. The main instruments used in the three studies:

A number of the well-known vocabulary measurements that are used to assess EFL learners' lexicon were reviewed earlier in section 2.4.4. The following measures were chosen for the current investigations (see the discussion in the following section for

justifications). The measures used were three vocabulary tests, a 5-point scale questionnaire, and the Lexical Frequency Profile (LFP) of learners' written production. The three vocabulary tests plus the 5-point scale questionnaire were used with all the participants, and the LFP analysis was used with a representative population of participants.

The four vocabulary measures used in *study one* (the first study) and *study two* (the second study) are:

a) Learners' receptive knowledge of the AWL was checked by the AWL test part of (Schmitt et al. 2001)

b) Learners' productive knowledge of the AWL was checked by the paper and pen CATSS, the AWL part items, (Batia Laufer, personal communication)

c) Learners' general vocabulary size was checked by XK-lex

d) Learners' ability to use, and frequency of use of the AWL in free writing was checked by the *Vocabprofiler*.

*Study Three* in Chapter 6 used:

a) Likert 5-point scale questionnaire to report feedback from learners.

b) Open-ended questionnaires to report feedback from teachers.

#### *3.4.1.1. The vocabulary levels test VLT*

The vocabulary level test VLT was reviewed in section 2.4.4. However, only the AWL test items' section of the test was used from this four frequency levels test (2000, 3000, 5000, 10,000 and AWL). The academic vocabulary section attempts to estimate how many of the 570 words in the AWL are known (Schmitt et al. (2001), and this section

could be used apart from the rest of the test items to assess the academic vocabulary (Norbert Schmitt, personal communication 2009).

Schmitt et al. (2001) compiled many versions of the VLT, and versions A, and B were used in *studies one* and *two*. The test contains thirty items divided into 10 sections. Each of the three items is given six options of possible answers (see the example in Table 3.3). Learners were asked to put the number of the appropriate word on the left beside its definition/synonym on the right.

**Table 3.3 Sample from AWL section of the VLT (version A) (Schmitt et al., 2001)**

| **1** | | **2** | |
|---|---|---|---|
| 1. benefit | | 1. achieve | |
| 2. labour | ___ work. | 2. conceive | ___ change. |
| 3. percent | ___ part of 100. | 3. grant | ___ connect together. |
| 4. principle | ___ general idea used to guide | 4. link | ___ finish successfully. |
| 5. source | one's actions. | 5. modify | |
| 6. survey | | 6. offset | |

The VLT is used to test the learners' receptive knowledge (recognition) of the AWL.

### *3.4.1.2. CATSS*

CATSS, or the Computer Adaptive Test of Size and Strength, was compiled by Laufer & Goldstein (2004) and Laufer et al., (2004). It samples some lexical items at all frequency levels, and tests four levels of knowledge of the same lexical item. The four levels that it tests are: *Active recall, Passive recall, Active recognition, and Passive recognition.* The four types were reviewed and examples given in section 2.4.4. As previously mentioned, the test is available in a computerized version that works on the basis of a certain formula that goes into a *hierarchy*. However, in the current research

(*studies one* and *two*), only the AWL section of this test was used. A paper and pen version of CATSS (Batia Laufer, personal communication[3]) was used in the study as it serves better logistically, and there was no necessity for the *hierarchy* that the test operates. Concerning the four levels that CATSS assesses, only the deepest knowledge test was applied in this thesis (studies *one* and *two*) which is the *Active recall test.* CATSS here asks the student to supply the word with the help of an initial letter to avoid non-target words. Below is an example:

**Table 3.4 Sample from the AWL section of CATSS (Laufer & Goldstein, 2004, Laufer, personal communication)**

| | |
|---|---|
| 1. general idea used to guide one's actions | p_____ |
| 2. moving from one country to another | m_____ |
| 3. acting against the law | v_____ |

In the current study, this test aimed to assess the productive knowledge of learners' academic vocabulary. The original test was criticized for the fact that it gives a hint to the learners by providing initial letters of the tested items. In order to counter this criticism, in this study, CATSS is combined with another uncontrolled productive writing test (see discussion in section 3.4.1.4. below).

### *3.4.1.3. The XK-lex*

*The XK-lex* (Meara & Milton, 2006) is used to assess the overall lexical competence of advanced learners at their last stage of instructed language learning. XK-lex is a yes/no test format that asks learners to put a tick beside the words they recognize. It samples words containing some *non-words* (or *pseudowords*) to control for any possible

[3] In 2009

overestimation by test takers. Besides measuring the vocabulary size of learners, XK-lex was used also to identify the correlation between the learners' overall lexical abilities and their knowledge of the academic words at the advanced level.

As mentioned in section 2.4.4., this test was used for many reasons. It covers more sample size of words than X-lex, which only covers the first 5000. Also, it is a paper and pen test which works better logistically for the study than the X-lex and EVST. It does not demand much time for administration as other more in-depth tests do, and, therefore interrupted less teaching time for the students. The test was validated by Masrai (2009). As it is a *tick box* test (see the example), it is thought to give a broader estimate of learners' level of vocabulary size than other vocabulary size tests that use a *multiple choice* or *productive test* format (e.g. the VLT or the VSL see section 2.4.4). However, in this research (studies *one* and *two*), all participants took two versions of the test which, therefore, could give a more stable score of level size. Moreover, although the VLT and VSL tests assess the lexical items more confidently than Xk-lex, because of the way they work (score based on ticking beside known words), this research has applied extra measures that test vocabulary knowledge.

**Table 3.5 Sample of XK-Lex (version A) (Masrai, 2009)**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **New** | | **Commerce** | | **Organise** | | **Accuse** | |
| **Gummer** | | **Tindle** | | **Wookey** | | **Candish** | |
| **Word** | | **Dust** | | **Fountain** | | **Tend** | |
| **Near** | | **Nonsense** | | **Movement** | | **Landing** | |
| **Peace** | | **Fond** | | **Likely** | | **Volume** | |
| **Produce** | | **Sweat** | | **Provide** | | **Tube** | |

The test presents 120 words (100 real words and 20 pseudowords) that represent the most common 10,000 words in English (each 10 words represent 1000 words). In working out the raw score for participants, each real word chosen as 'known' counts as 100 marks,

and in the final score 500 marks are deducted for each unreal word chosen as 'known'. See versions A and B in the Appendices. See also section 3.4.3. for how the test was marked.

### *3.4.1.4. Vocabprofiler: checking the lexical richness of the AWL in free writing*

The vocabulary profiler *vocabprofiler* (Cobb, 2008) is a web tool that checks the Lexical Frequency Profile (LFP) of learners' written production. It calculates the words shown in the written text and assembles (categorizes) them depending on their frequency band, in order to state the proportion of the GSL, AWL, ESP words or other words, from learners' written passages. The AWL frequency percentage and the count of *Types* are our two main concerns in the analyses for this research.

Table 3.6. below shows an example of how *vocabprofiler* analyses the written text of 186 words of a typical academic writing task;

**Table 3.6 Words categorized using the online *vocabprofiler*[4] out of the 186 words passage: (definitions of families, types and tokens are found in 2.4.4)**

| | Families | Types | Tokens | Percent |
|---|---|---|---|---|
| K1 Words (1-1000): | 86 | 97 | 154 | 82.80% |
| K2 Words (1001-2000): | 15 | 16 | 16 | 8.60% |
| AWL Words (570): | 8 | 8 | 8 | 4.30% |
| Off-List Words: | ? | 6 | 8 | 4.30% |
| Total | 109+? | 127 | 186 | 100% |

While using this instrument in the thesis, the main focus was on how often the AWL words appeared in learners' free writing tasks. i.e. how rich the text was in terms of AWL words. Some researchers have concerns about 'free' and 'uncontrolled' productive measures which could cause difficulties for the tester as learners may produce unpredictable results (as discussed in section 2.4.3). That is why CATSS could work

[4] Retrieved at http://www.lextutor.ca/vp/eng/

better in this situation. However, CATSS might still be criticized for giving some hints about the answers by providing the initial letters (as mentioned above).

In the current study, both tests have been used to assess the learners' knowledge of the AWL. Regarding the 'difficulty' and the possibility of 'unpredictable' responses of free uncontrolled production assessment argued earlier, for the present two studies this might not be considered as a serious concern as it focuses only on how many of the academic words appear in learners' texts, and does not predict a particular sample of academic words. In other words, the study only counts the proportion of correct responses of the academic words even if the responses of student A, for example, are different from student B, as long as the words used are within this list of academic words.

Only a representative sample of students was involved in this measure.

### 3.4.2. Reasons for choosing the multiple instruments used in this thesis.

It has been extensively argued that multiple measure of vocabulary is crucial because of the nature of vocabulary knowledge and the non-clearly defined relationship between size and depth (see section 2.3.1. about *word knowledge*, and 2.3.1.2. about the *relationship between size and depth*). Scholars emphasize this more in the case of longitudinal studies and/or experiments that explores vocabulary growth (Nation, 2007b; Schmitt, 2010a). Thus, the present study applies different measures to test the academic words of advanced learners receptively and productively. It would be even more important in this study to apply multiple testing as it involves vocabulary leaners from different contexts, since an assessment of a single aspect of word knowledge may not provide accurate results regarding word gains, nor variations between learners' groups.

A number of vocabulary tests have been reviewed but some were thought to be unsuitable for this type of study due to conceptual, logistic or administrative reasons. The following is a brief rationalization for the use of the tests finally decided on:

- Since this study targets the AWL, only relatively up-to-date tests contain the AWL that was compiled by Coxhead in (2000). Sampling a representative range of words from the AWL and adapting some of the old vocabulary tests might not be suitable at this stage, and would affect the general analysis of the results of this study (O'Sullivan, 2010, personal communication). Also, as argued earlier (see 2.4.3), in experimental studies, it is recommended to use well-known and accepted tests.
- Since the study planned to use multiple measures, it ensured avoidance of long and time-consuming tests; i.e. if longer tests were used, it would have interrupted schooling time more, and this might have cut down on the number of measures used.
- The study excluded any test that would require students to use computers in order to have similar targeted words. Having similar targeted words for all participants might not be possible with some computerized tests that run formulas to form representing sample words for each test taker as discussed earlier. For logistical reasons, pencil and paper tests would guarantee a larger sample size of participants.
- Some of the tests do not provide larger representative test items of academic vocabulary, therefore, they were also excluded.
- The study combines *CATSS* and *Vocabprofiler* as two writing productive measures. CATSS is a controlled production test that asks students to write the targeted words with the help of defining synonyms and an initial letter to limit the choices.

Vocabprofiler calculates how frequently academic words are used in free writing tasks.

- As mentioned earlier in section 2.4.3., Schmitt (2010a) argues that some versions of the same test might not produce statistically equivalent scores. This risk would become even higher since this study includes only a smaller part of the test (the AWL part). To avoid the complication of same/different version issue, the present study has in fact applied the two situations of assessment tools. It uses different versions in the receptive knowledge test (the VLT) in the pretest and posttest, and used the same sheet of CATSS in pretest and posttest.
- Xk-lex was used to measure the overall size level as it is easier to administer, less time consuming, and represents larger vocabulary bands (above 5000) unlike X-lex (AlSaif, 2011; Masrai, 2009).

Table 3.7 below summarizes the tests that were used in the study.

**Table 3.7 The tests used in this study:**

| Type of test | What is meant to test |
|---|---|
| Version A & B of the AWL part of the *VLT* (N. Schmitt et al., 2001) | Pre-and post-testing of learners' receptive knowledge. |
| The AWL paper version of *CATSS* (Batia Laufer, personal communication). | Pre-and post-testing of learners' productive knowledge. |
| The *vocabprofiler* (Cobb, 2008) | Exploring lexical richness of the AWL in learner' free writing tasks |
| Version A & B of *XK-lex* (Meara & Milton, 2006) | Measuring overall vocabulary size of learners |

Concerning the discussions in section 2.4.3 regarding the dimensions of vocabulary assessment, the present thesis uses popular tests that come in many of these formats (selective, comprehensive free writing, and context-dependent and independent), as

discussed in section 2.4.4. Regarding *comprehensive* assessment, Schmitt (2010a) comments that 'free' output poses difficulties for the tester as the learners' answers and responses will be uncontrolled and might be unexpected. However, although the present thesis has included 'free' writing among the measures used for the experiment, unexpected responses should not be a serious concern for the following reasons:

Firstly, the written tasks have been carefully designed to allow fewer chances of unexpected responses. The fact that the second study involves learners from two different learning groups (i.e. receiving the academic vocabulary in two different ways), the study expects some general results that would meet, at least, the minimum objectives of this study. In other words, the complications of 'free' output tests mentioned earlier might not seriously affect the results of this study because it involved two different learning groups. Related to this, the present study included multiple measures apart from 'free writing' which would support the analysis of the potential results. Lastly, the inclusion of this type of measurement in the study is for a specific purpose that should not be affected by the drawbacks of 'free uncontrolled' output. In fact, the analysis of the studies in this thesis involves only exploring how frequent academic vocabulary would appear in their writing. The thesis does not expect students to produce exactly the same words it meant them to produce, but rather a range of words within the academic word list.

Concerning the memory factor and the pre-test/post-test issues, the study considers the following; first, in two of its measures, it uses two different versions of the same tests that have been widely validated in literature. With the controlled productive test, it uses the same version for the pretest and posttest to avoid the possible complications of applying two different versions of the test. Schmitt (2010a) mentioned that tests that involve

different items in the pre-test/post-test are likely to result in unknown variation. However, when using the same tests for the two occasions, he recommended using some techniques to cut out the memory effect. He suggested giving the participants a cognitively challenging task immediately after the test administration, in order to get them thinking about something or to distract them. This study has applied this as learners took one test after another immediately. Lastly, when using two different versions, Schmitt (2010a) recommended involving larger groups of participants, as individuals might produce scores that are not statistically equivalent. This study follows this recommendation.

### 3.4.3. A description of the participants in general, and in their respective studies

Further details about all participants of the three studies of this thesis are mentioned in Chapters 4, 5 and 6 respectively. However, since the three studies were related to each other, and learners' participation overlaps between *study three* and studies *one* and *two*, the general information about the participants in all three studies is presented in Table 3.8 below.

**Table 3.8 statistics of all participants of three studies;**

| | Major | | | | PG, or UnderG | |
|---|---|---|---|---|---|---|
| Students' Groups [a] | Humanities | Medicine | Science | Pre-study | PG | UG |
| SUni 1 | - | - | - | 50 | 42 | 6 |
| RUni 2 | - | - | - | 44 | 41 | 2 |
| SA-DG | 33 | 8 | - | - | 41 | - |
| SA-NnDG | 27 | 11 | 1 | - | 39 | - |
| Total 174 | 60 | 19 | 1 | 94 | 163 | 8 |

[a] study one: two universities involved (SUni1, RUni2), study two: two Saudi groups at one university but were involved two groups, and study three included all participants as explained below.

The first study targeted advanced students at their last stage of English study (i.e. presessional courses), which come just before the start of their graduate or postgraduate

major. The study selected learners from two universities to limit any learning factors that might occur in applying the study to a single location (see the discussion in 4.2.1.). The two universities were expected to run similar presessional programmes as both of them were selected from the 20 top-ranking UK universities. 101 advanced learners with a proficiency equal to IELTS 5.5 and above from two universities participated in the study, which is the condition of entry onto presessional programmes. Further discussion of learners and the presessional programme is given in section 4.2.1.

The second study targeted advanced students at Preparatory (or foundation) Year PY at King Saud University KSU in Riyadh, Saudi Arabia. Learners at the PY programme are assigned into 6 levels according to their proficiency (where level 6 equals high achievers), and participants at this study were from level 5. There were not many students at level 6 at the time of the study; therefore, learners from level 5 were selected. 87 male students participated in the second study. As mentioned in the procedure (in section 3.4.), participants were from two learning groups. There were 44 learners from the treatment group or direct group DG, and 43 from the control group or none direct group (NnDG henceforth). Further details about the PY programme and the participants are presented in Section 5.2.1.

*Study three*, about *feedback and learning beliefs of the AWL*, asked the participants of *study one* and *study two* and their teachers to participate in the survey. As mentioned earlier, *study three* was dealt with as a separate study as it explores more research questions regarding learning beliefs and perceptions, apart from the feedback about learning the AWL (see section 6.1.). The participants included most of the learners from the previous two studies who had agreed to fill in the questionnaires. They consisted of at

least 80% of the population which took part in *studies one* and *two* (see section 6.2.2. for further details). The teachers were 15 of the language tutors from two UK universities and from KSU University in Riyadh.

### 3.4.4. The general procedure of the analysis of the data

*Scoring and coding the data:* All the data from the three tests the VLT A and B, CATSS and XK-lex A and B were itemized and transformed into numerical codes and put into *Excel* tables to allow easy conversion into SPSS. Each correct response of each test was given '1' and each incorrect or missing answer was given '0'. However each test has a different marking/coding scheme depending on the design and content of the test.

Responses of the *VLT* recognition multiple-choice items were scored (1) for correct answers and (0) for incorrect or missing responses. To mark CATSS results, since it was a productive test and learners were expected to have some spelling errors, all the papers were reviewed manually for a provisional analysis before the item entry of the data and the decision of the marking scheme was set. Two types of spelling errors found in learners' answer sheets included: *minor error*, where missing or misspelling vowels occurred, or *major error*, where the word might be read with difficulty, or was found to be confused with another word. Only the minor errors were accepted as correct answers (given 1). Others were given '0'.

Regarding the vocabulary size test (XK-lex), as mentioned before, the test presents 120 items (100 real words and 20 nonsense words), and asks respondents to tick beside the words they think they know. So the possible responses are:

A- Tick beside the real word as known (correct), B- leave the nonsense word (correct), C- leave the real word (incorrect) or D- claim the *nonsense invented* word as known

(incorrect). In this study only, the coding was '1' for the correct answers in items A and B above and '0' for the other options.

Concerning the *vocabprofiler* and analysing the lexical richness of the AWL in the writing sample, *studies one* and *two* follow the same analysis scheme of the software set by Cobb (2008). However, due to the focus of this study, only the figures and frequencies related to the AWL were the focus in the analysis. Besides, since the written data was submitted on a paper and pen basis, and some participants were likely to have spelling mistakes, the unclear or the barely decipherable words from the written tasks were shown to a native English speaking colleague as a validator. The purpose was to have a fixed procedure in the scoring scheme. The validator was simply asked:

| ID | Passage | Paragraph | line | Word | *What do you think the student meant by this word?* |
|---|---|---|---|---|---|
| 1 | 2 | 4 | 1 | …… | |

The writing tasks are all typed into MS Word processer to be analysed with the *vocabprofiler*.

Two analyses of the written data were run: one that included the whole length of the text (all word count) of each writer, and the other included only the first 200 words (based on the shortest composition written), as will be illustrated in the results below.

Laufer and Nation (1995) claimed that the lexical frequency profile presented similar results for pieces of writing by the same writer. Their study was criticized by Meara (2005) for the fact that it does not reliably distinguish between learner groups at different levels of proficiency or vocabulary size. The present study, however, only focuses on a simple count and frequency of the AWL in participants' writing. Besides, learners came from a similar proficiency levels and their vocabulary size was also assessed.

Since the study involved collection of the writings from a group of students rather than just a single participant, this ensures a clearer picture of any occurrences of development over a period of time. It also allows more confident generalization of the results regarding the variations between the pre- and post- writing tasks.

*The data analysis*:

The data collected was analysed according to the aims and objectives set for each study.

A number of data analysis steps were carried out in order to produce the following results:

- General means, maximum and minimum scores, and standard deviation of the three tests (VLT, CATSS and XK-lex)
- Correlations between scores of different tests.
- Rate of increase/decrease (difference) between pre- and post-test.
- Where applicable in one of the studies: difference between gains of the two learning groups.
- Correlation between the tests and the ability to produce academic vocabulary in free writing (*vocabprofiler).*

The SPSS statistical package was mainly used for the data analysis to produce the descriptive and correlation results mentioned above. Below are the five steps of the data analysis of *studies one* and *two*. However, since in *study two* (the second study), two learning groups were involved, a sixth step was added to the procedure of the analysis as will be seen shortly:

- First, the reliability test was performed using SPSS. All the items of tests of all the groups have been checked for reliability using Cronbach's Alpha.

- Second, since the tests vary in terms of the number of given items, they were all transformed into percentages for clearer and easier display of correlations. To illustrate, the different measures have different maximum scores (i.e. CATSS=30, VLT=30, XK-lex =120, and *Vocabprofiler* = Count of words frequency within text). It was decided that if all measured scores should be transformed into percentages, it would be easier for the analysis and clearer for the reader as it would make the scores more clearly equivalent; see Table 3.9 below:

**Table 3.9 Measures maximum scores transformed to percentage:**

| | Instrument | N Items (maximum score) | Transformed to |
|---|---|---|---|
| 1 | The VLT | 30 | %100 (=sum/30*100) |
| 2 | *CATSS* | 30 | %100 (=sum/30*100) |
| 3 | XK-lex | 120 | %100 (=sum/120*100) |
| 4 | Vocabprofiler | Word Count of AWL | %100 (=sum/WCA*100) |

- Third: a descriptive analysis was made which shows the overall score results, means, maximum and minimums and standard deviation. T-test or one of its equivalents (depending on the normality of the data) was used to determine the change between pre and post scores of each measure.
- Fourth: the statistical method two-way repeated measure ANOVA (analysis of variance) was employed to determine the relationship between all measures and to compare the scores of the pre and in the post occasions to retrieve the difference. Two-way ANOVA is employed in the second study to retrieve the Pre-Post difference in each measure of each group.

- *The Correlation Coefficient* was used to examine the relation between pre and post scores of each measure. Figure 3.1 below summarizes the analysis procedure of the first study:

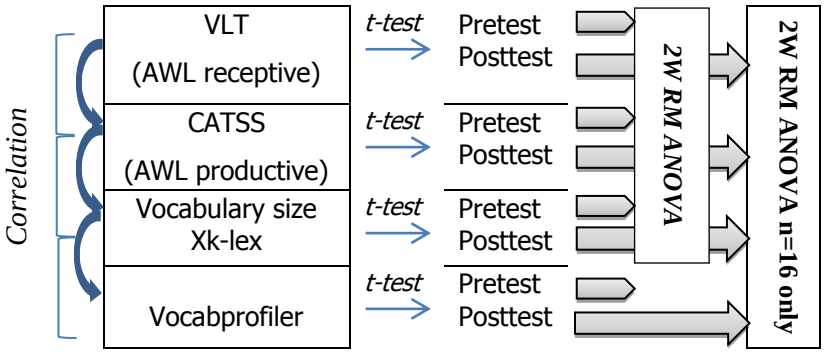



**Figure 3-1: The relation and correlation applied between all measures:**

- Sixth: as mentioned earlier, this step was only applicable for *study two* (the second study) as it had learners from two taught groups (DG and NnDG) who experienced different methods of learning and exposure to the AWL. An extra process of analysis was performed here. A two-way mixed design *MANOVA* was run to compare all scores in the two learning conditions to determine if the means were statistically different. The t-test (or its nonparametric equivalents) was performed to compare the means of the two groups in each test type.

The figure below (Figure 3.2) summarizes the procedures of the *second study*:

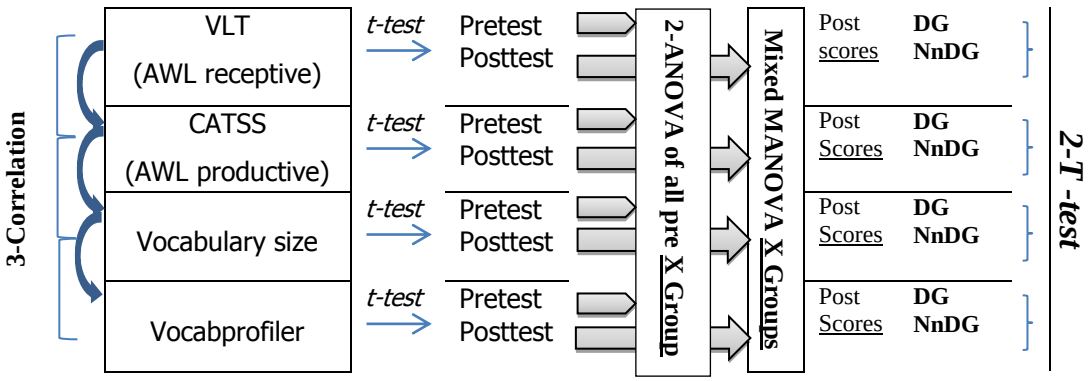



**Figure 3-2: The relation and correlation applied between all measures:**

The vocabulary profiler follows the same analysis scheme set by Cobb (2008) as mentioned earlier in this section.

*Study three:* Feedback of learners and teachers;

The SPSS statistical package was used to analyse the responses to reveal the mean scores with respect to the research aims and questions. Correlations were run to demonstrate any factors related to the background information of respondents. NVivo was used for coding and reporting the responses of open ended questions. Further details are in Chapter 6.

## 3.5. Summary of organization of the three studies of this thesis

This chapter has discussed the rationale of the thesis with respect to each study of the paper, and defined the research questions that the different studies explore. It also described the methodology, the participants and procedures of research and data analysis from a general perspective.

In the following chapters, Chapter 4 investigates the learning of the academic word list on typical presessional courses in the UK, Chapter 5 investigates the learning of the academic word list by learners receiving two different methods of teaching on an advanced intensive English programme, and Chapter 6 analyses learner and teacher feedback on learning the academic word list. Each chapter explains with further discussions the specific rationale of each study, and gives detailed description of the methodology. Each chapter starts with an introduction and general aims of the study, and lists a number of specific research objectives based on the literature review. It then lists the research questions of the study, and how they will be explored.

Each chapter follows with full descriptions of the research design and methodology of the study. This includes detailed information about participants, the instruments used, the procedure of the research and how the data was analyzed. This is followed by presentation of the research findings, with approximate interpretation of how the overall result reads. The chapters conclude by summarizing each individual study.

# 4. Chapter Four

## The First Study:

## Investigating the Learning of the Academic Word List on typical presessional courses in the UK.

## 4.1. Introduction and purpose of this study.

This study explores the lexical learning needs of advanced EFL students willing to study at university level in an English-speaking country. It also investigates how the academic exposure within advanced EFL programmes allows the learning of the academic words.

It was argued in Chapter 2 that vocabulary is very important for the development of language skills, and that vocabulary size contributes significantly to learners' competence. It has also been argued that knowledge of the academic vocabulary list (Coxhead 2000) has a positive impact on learners' language proficiency both receptively and productively. However, as previously discussed, there are some research gaps regarding many issues in this area. Only a few studies that have measured the vocabulary size of advanced learners willing to study at the academic level are available, and many of these targeted a specific population of learners (see for example Schmitt's review, in 2008a). Besides, the fact that different vocabulary size tests and test formats were used to check learners' lexical level (AlAkloby, 2001; AlHazemi, 1993; AlHomoud, 2007; AlNujaidi, 2003), and that some of these tests have been recently complied (Cobb, 2008; Laufer & Goldstein, 2004; Masrai, 2009), suggest that more investigation into advanced learners' overall lexical knowledge is needed.

Apart from the general vocabulary size of learners, there seems to be some vagueness about the actual level of advanced learners regarding knowledge of the academic word list; i.e. only a few studies have explored the abilities of advanced learners regarding the AWL (Chui, 2006). Furthermore, with the complex nature of knowledge of words and how lexical items are acquired, it is important to explore different angles of learners' abilities regarding vocabulary. This would be useful in terms of checking learning needs

and, therefore, allow better decisions for classroom planning and pedagogy in general. Finally, since the advanced EFL learners are commonly perceived among educators to be capable to learn words incidentally, it is important to investigate how likely the AWL is learned from the academic exposure within advanced EFL courses.

With respect to the discussion above, this study aims to address the following research question:

> **RQ 1-** How does the depth and breadth of academic vocabulary (as measured by the AWL) of learners at an advanced stage of their pre-university studies change during their preparatory programme?

As discussed in section 3.3, an additional sub-research query emerged, and a number of research hypotheses were put forward. In order to respond to the research questions and research hypotheses, this study attempts to:

1. Determine the general vocabulary size of advanced learners at the last stage of their English preparation programme before moving on to university level.
2. Identify how many words of the AWL learners at this stage of advanced language learning know.
3. Check the productive abilities of learners regarding this vocabulary genre using extra multiple measures to test further aspects of mastery due to the complex nature of vocabulary knowledge.
4. Investigate the relationship between the general vocabulary size of learners and their ability to recognize and use the AWL.
5. Investigate the lexical richness of the AWL in learners' advanced writing.

## 4.2. Method

### 4.2.1. Participants.

The participants of the first study are advanced EFL learners at the final stage of their language learning before accessing academic courses at UK universities. Learners from two universities were involved in this study in order to present more generalizable results and present a more confident analysis of the data about the possible factors affecting learning. Conducting the study with similar learners studying at two universities (instead of one) would allow for a broader generalization regarding the main queries of the study, as it would eliminate possible factors that might be thought to have had an effect on learners' results e.g. the university hosting the study, its policy or curriculum, students, teachers. Besides, this study targeted students from two of the top twenty universities in the UK in order to ensure relatively similar learning conditions and try to target as similar a level of student as possible.

Many universities run EAP and presessional courses to prepare EFL learners before they study their academic majors. EAP courses are run all year round in many universities, but many presessional courses are run in the summer and usually admit only the most advanced and most able students. Usually the students are highly motivated, especially those who have not yet reached the English proficiency level required by their individual departments, often ranging between 6.5 and 7.0 in the IELTS test, 575 in TOEFL, or 90 in iBT. If students are a few points short of the required mark (for example 5.5 or 6.0) they are assigned to presessional courses of different durations to raise their existing band or level. It is generally expected that each (0.5) is equal to 4 to 5 weeks of coursework.

The participants in this study were from courses of 10 weeks' duration, so are likely to have been at an IELTS level of approximately 5.5 or 6 on arrival.

101 students from two universities were involved in the study; 50 from university one (henceforth: SUni 1) and 51 from university two (henceforth: RUni 2). Table 4.1. below shows more detailed information about the participants, information obtained from the participants' questionnaires.

**Table 4.1 Overview of information about the learners participating in this study:**

| | | PG, or UnderG | | Notes |
|---|---|---|---|---|
| | | PG | UG | -Participants from: China, Taiwan Saudi, Oman, Jordan and Spain. |
| SUni 1 | | 44 | 6 | |
| RUni 2 | | 49 | 2 | -Males and females. |
| Total | 101 | 93 | 8 | -Ages ranged from 22 to 35 |

*PG: postgraduate, UnderG: undergraduate*

Learners on presessional courses are expected to attend between 4 to 5 hours of classes a day (Monday to Friday), plus engage in about two hours of independent study. Most of the curriculum materials aim to enhance the academic skills required for students' majors. They practise academic writing: writing compositions similar to IELTS tasks or more in-depth essays related to their subjects, or summarize research papers or reports from their academic areas. From personal experience, learners are often assigned to read academic articles and learn to criticize texts. Academic listening is part of these courses e.g. listening to lectures presented by a guest, and taking notes, and finally they work on academic speaking, developing skills for giving presentations or engaging in seminar discussions.

In the two schools, vocabulary was mostly learned incidentally through exposure to the academic skills classes (further discussion on this can be found in the results from the

teacher and learner survey in Chapter 6). By the end of the 10th week, learners are expected to have been exposed to about 225 hours of academic English (normally 3 sessions a day, 90 minutes each).

### 4.2.2. The instruments used

The instruments used in this study were the three tests described in Chapter 3 plus the vocabprofiler (see section 3.4.1.). It has been argued several times in this thesis that multiple measuring of vocabulary is essential due to the incremental process of vocabulary learning and word knowledge. This, as discussed earlier in the chapter (see section 4.1), reinforces the use of multiple tests in exploring the lexical capabilities of learners for research or pedagogical purposes.

The multiple tests used were:

- The vocabulary level test VLT (Schmitt, Schmitt, & Clapham (2001),
- CATSS or the Computer Adaptive Test of Size, a pencil and paper version, (Batia Laufer, personal communication)
- The XK-lex (Meara & Milton, 2006), and
- The vocabulary profiler (vocabprofiler) (Cobb, 2008).

Table 4.2. below summarizes of the instruments used and what they are meant to assess, the participants, and the timing of tests;

**Table 4.2 A summary of the tests and procedure:**

| Instrument | What is meant to be tested | Participants | Time |
|---|---|---|---|
| Version **A** of the VLT | Receptive knowledge of the AWL. | All students | Week **2** of the course |
| Version **B** of the VLT | Receptive knowledge of the AWL. | All students | Week **10** of the course |
| Adopted version of *CATSS* | productive knowledge of the AWL | All students | Weeks 2 and **10** of the course |
| Writing Task **A** and vocabprofiler | Lexical richness of the AWL in writing | 8 students | Mid Term |
| Writing Task **B** and vocabprofiler | Lexical richness of the AWL in writing | 8 students | Final |

The writing tasks that some participants completed for the *vocabprofiler* analysis are discussed in the following section.

## 4.3. Procedure

The three tests (the AWL sections of; the VLT and CATSS, and the level size test; XK-lex) were administered to all participants from the two universities (*SUni 1* and *RUni 2*). However, for logistical reasons, in *SUni 1*, all students took their tests at the same time in a big lecture theatre. The teachers helped in the administration of the tests. In *RUni 2*, the participants took the tests in their individual classrooms (n 12 to 15) after the procedure and test instructions had been fully explained to their teachers. All learners were encouraged to write their names in all of the different test sheets they were given to allow further comparison analysis between different measures. They were aware of their rights, specifically that the data would be kept confidential, and that the purpose for writing their names was only to match the relevant test sheets.

Concerning the *lexical richness* of the academic words in learners' writing and the use of *vocabprofiler*, only some of the students were involved in this type of measure.

All participants of *SUni 1* and *RUni 2* took these tests twice on their presessional programmes. They were tested with the pre-test at the beginning of the course and post-test in the last week before the course ended. In the tests; VLT and XK-lex, versions *A* were used for the pre-test and versions *B* were used for the post-test. However, the same adopted version of CATSS was used twice in the pre-test and post-test. The fourth measurement, which involves checking the lexical richness of the AWL in learners' writing using the *vocabprofiler*, was carried out by collecting two writing tasks from a representative number of students; one at the beginning and one at the end. It was ensured

that the two academic writing tasks were similar in their topics in order to minimize any factor that would affect choice of vocabulary resulting from the topic. The pre- and post-writing tasks that were given to the participants are stated below. Summaries of the instruments, participants, and timings of the assessments are in Table 4.2.

**Task A:** Write at least 250 words on the following topic:
"What is the main purpose of education? Give reasons for your viewpoint".

**Task B**: Write at least 350 words on the following topic:
'What are the advantages and disadvantages of being self-employed, compared to working for a large organization? Why are some people more suited to being self-employed than others?"

## 4.4. Scoring and coding procedures and data analysis

Detailed descriptions of scoring and coding procedure of the data are discussed in section 3.4.3. Generally, the responses of participants in all of the three measures were itemized, transformed into numerical codes, and input into *Excel* to be easily processed with the SPSS statistical package. The final scores were calculated by counting how many correct answers (i.e. how many item have a value of '1') and excluding the missing or the incorrect responses (item value '0').

Each test (plus *the vocabprofiler*) has a particular scoring procedure as explained in the Methodology Chapter in section 3.4.5.

Concerning the data analysis, only the first five procedure steps mentioned in section 3.4.3. were performed in this study: a) the reliability test, b) the transformation of all scores of the different tests into percentages to make correlation analysis results more comparable and clear (see the discussion in 3.4.3.), c) the descriptive results of mean scores, minimum and maximums and standard deviation, d) the *ANOVA* to compare the four pre scores and the four post scores, and e) the *t-test* and *the Correlation Coefficient*

between the measures, as well as the *vocabprofiler*. As mentioned in section 3.4.3 those procedures aim to demonstrate:

- The overall score results of all tests, means, minimums and maximums and standard deviation
- The correlation between the measures, as well as the effect of test scores on using the AWL in free writing
- The significance of increase or decrease of scores and the degree of the development (pre/post change).

The sixth step of the data analysis procedures was excluded here because it is only applicable to the main study.

## 4.5. Results and discussion

The main research question and hypotheses involve exploring the following:

- Learners' improvement regarding the receptive or recognition level of the AWL.
- Their improvement regarding the ability to produce the AWL in a productive controlled test.
- Their improvement regarding the ability to produce the AWL freely in writing.

As discussed in section 3.3, the main research question suggested a further additional sub-research query. This additional question involves investigating:

- The relationship between the different measures.
- Learners' of vocabulary level.

The above items entail determining the initial levels of learners at the beginning of their programmes (pre-test scores), their knowledge at the end of the programme (pre-test scores), and the significance of change (i.e. improvement) in all measures.

The results concerning the above points are discussed in turn below. Note that, for the sake of organization, the level of the general vocabulary size that concern the sub-research query will be dealt with before addressing the relationship between all measures.

### 4.5.1. Advanced learners' receptive knowledge of the academic words.

#### *4.5.1.1. Reliability and normality*

As described in the procedures in section 3.4.3., the data of the VLT academic recognition test of all participants of the two universities was examined to ensure that the data is reliable using *Cronbach's Alpha* test of reliability. Also, the items of all participants were checked for normality of distribution using the *Shapiro-Wilk* test to confirm which statistical analysis was to be performed.

Table 4.3. below shows the figures of reliability and normality distribution:

**Table 4.3 Tests of Normality and reliability in SUni1 and RUni2 and both universities**

| School | Rec test (VLT) | Cronbach's Alpha | Shapiro-Wilk | | | N |
|---|---|---|---|---|---|---|
| | | | Skewness | Kurtosis | sig | |
| SUni1 | Pre | .816* | -1.199 | 1.739 | .100 | |
| | Post | .823* | -2.108 | 4.992 | .007 | 51 |
| RUni2 | Pre | .832* | -.438 | -.548 | .001 | |
| | Post | .813* | -.750 | -.008 | .000 | 50 |
| 2Univs | Pre | .836* | -.730 | -.026 | .000 | |
| | Post | .830* | -1.242 | 1.062 | .000 | 101 |

* $P > 0.05$ abnormal

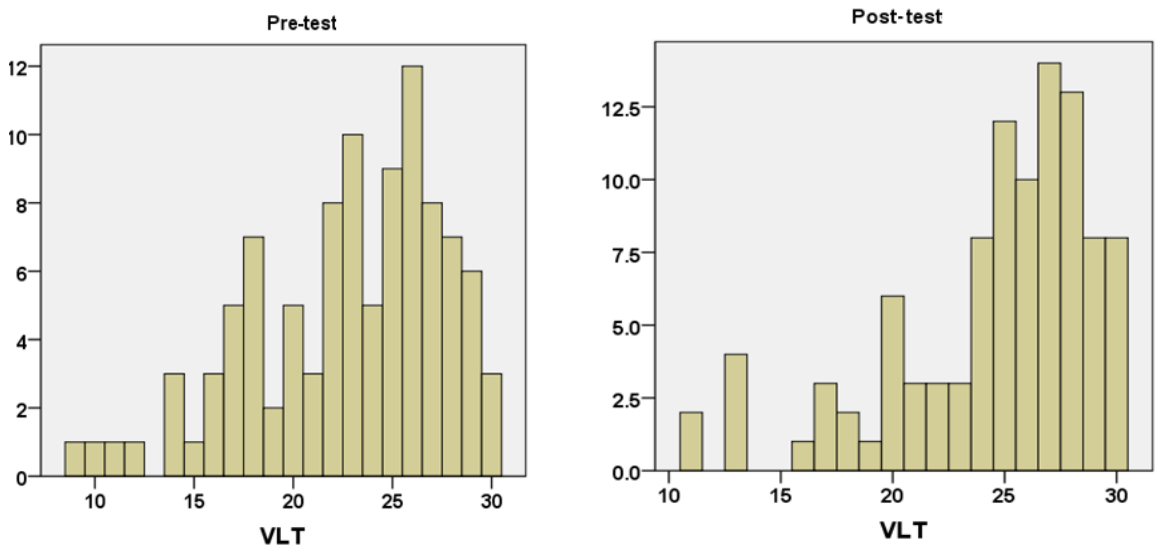



**Figure 4-1: Histograms of the AWL receptive test VLT**

The reliability of Cronbach's Alpha for the items of the VLT (receptive test) of the AWL was as high as (0.8). Concerning the normality of distribution of the data, the figures show that the scores of the *SUni1* have 1 acceptable figure of normality (.100) in the pre-test. However, in the post-test, all the data of *RUni2* and taking the data as a whole show that it is not normally distributed. The reason behind this could be the fact that there is a clear sign of skewness in the data towards the maximum score, and the minimum score in *RUni2* was about 15/30 as seen below. The histograms in *Figure 4.1* show further details of distribution and degree of skewness of the data of this measure. Those figures regarding normality of distribution suggested the use of non-parametric analysis of the data.

### *4.5.1.2. Descriptive statistics of the Receptive test of the AWL*

*The initial abilities of learners on presessionals:*

The scores of the pre-test show that advanced learners at the beginning of presessional programmes could recognize at least two thirds of the academic vocabulary that the VLT

represents. Their average score was 22.75/30. There was a moderate variation between students' scores at about ±4.6. Surprisingly, there were some lower achiever students who hardly reached a score of 9/30. This score could be considered unexpected with advanced EFL students at this stage of language study. According to the formula of Schmitt and AlHomoud (AlHomoud, 2007), this low score is equal to 171 words of the academic list.

*(Score ÷ 30 Maximum) x 570 AWL = estimated known words.*

Based on this formula, the estimated average number of academic words that all students were able to recognize was 433 words. It is clearly noticed that learners at *RUni1* had higher scores in this test.

Table 4.4 below shows the pre-test mean, the standard deviation, the minimum, and the maximum scores of the academic section of the VLT (receptive knowledge) of all participants; the purpose of the table is to show some descriptive data, and the statistical comparisons will follow.

**Table 4.4 Descriptive statistics of scores of SUni1 and RUni2, and both universities**

| School | Test | Mean /30 | ± SD | Minimum | Maximum | Range | *N* |
|---|---|---|---|---|---|---|---|
| SUni1 | RecPre | 21.314 | ± 4.8 | 10 | 29 | 19 | 51 |
| RUni2 | RecPre | 24.2 | ± 4.4 | 9 | 30 | 21 | 50 |
| 2Univs | RecPre | 22.75 | ± 4.8 | 9 | 30 | 21 | 101 |

*AWL receptive improvement during the presessional programme:*

Although the scores of the pre-test were relatively high (reaching the maximum score), learners at both universities have shown some improvement of receptive knowledge of the academic words. The brief descriptive results are given in Table 4.5, below and illustrated in Figure 4.2.

**Table 4.5 Basic descriptive statistics of means and standard deviation of both tests of Receptive test (VLT)**

| School | PreTest | ± SD | Post-test | ± SD | *N* |
|---|---|---|---|---|---|
| SUni1 | 21.314 | ± 4.8 | 23.314 | ± 4.8 | 51 |
| RUni2 | 24.2 | ± 4.4 | 25.8 | ± 3.9 | 50 |
| 2Univs | 22.75 | ± 4.8 | 24.55 | ± 4.5 | 101 |

Max = 30

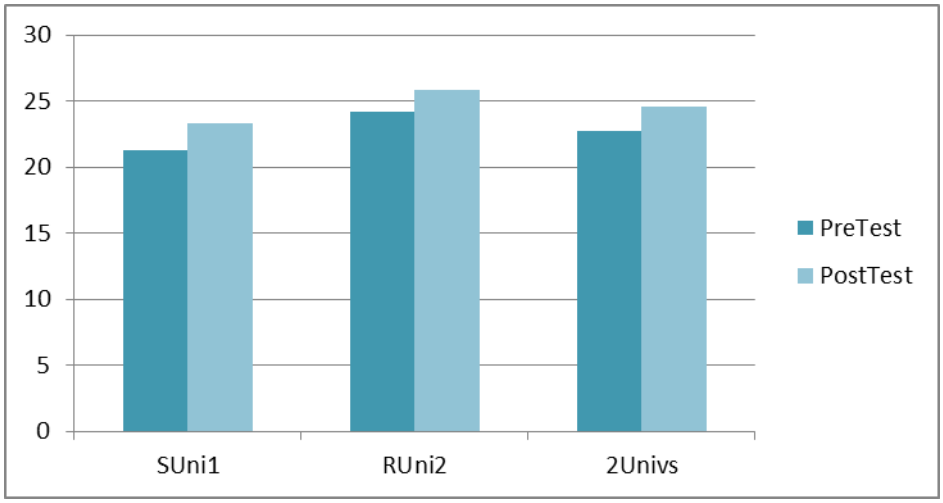



**Figure 4-2: Improvement of receptive knowledge of the AWL (VLT test)**

The descriptive statistics show some improvement in receptive knowledge. However, as expected, the increase was not large because learners had shown high scores initially. What could confirm this claim is that the learners of the lower group (*SUni1)* showed a slightly greater improvement than those learners at *RUni2*, who had higher maximum scores in the pretest.

Nonetheless, an analysis of comparisons was performed and confirmed the significance of increase among the students. The analysis was made by running a *Wilcoxon test* as an alternative due to the non-normal of distribution of the data reported earlier. The comparisons show significant pre-test/post-test differences at ($p<.001$).

Table 4.6 shows detailed figures of improvement levels in maximums, minimums, medians and difference;

**Table 4.6 Descriptive statistics of REC (VLT) scores of SUni1 and RUni2, and both universities combined**

| School | Test | Mean +-SD 30/30 | Dif | P Value | *Min* | *Max* | *Median 30/* | *N* |
|---|---|---|---|---|---|---|---|---|
| SUni1 | RecPre | 21.314± 4.8 | 2.0 | .000 | 10 | 29 | 22.00 | 51 |
| | RecPost | 23.314± 4.8 | | | 11 | 30 | 24.00 | |
| RUni2 | RecPre | 24.2 ± 4.4 | 1.6 | .001 | 9 | 30 | 25.00 | 50 |
| | RecPost | 25.8 ± 3.9 | | | 10 | 30 | 27.00 | |
| 2Univs | RecPre | 22.75 ± 4.8 | 1.8 | .000 | 9 | 30 | 23.00 | 101 |
| | RecPost | 24.55 ± 4.5 | | | 11 | 30 | 26.00 | |

It is no surprise that the figures show significant increases in the receptive knowledge of the academic words, since the basis of presessional courses is academic. However, the increase may be considered unsatisfactory. The rate of improvement between scores of both tests was (1.8) words. Adjusting the same formula to calculate the estimates, the rate appears to represent an increase of 34.2 words learned during the programme, at a rate of (.75) word a day.

### 4.5.2. Advanced learners' productive knowledge of the academic words.

#### *4.5.2.1. Reliability and normality*

The reliability and normality figures of the data of *CATSS* test of production of academic words of all participants were confirmed by *Cronbach's Alpha* test and the *Shapiro-Wilk* test. The figures show that *SUni1* had strong figures of reliability of items in both tests at (.750 and .849). Items of *RUni2* showed acceptable [5] figures of reliability in the pre-test and a strong reliability in the post-test (.569 and 809). The overall figures of the combined universities in both tests were (.663 and .826) respectively.

[5] Acceptable based on the population of the data and also the truncated values (Punch, 2005).

The normality figures showed normality of distribution in both tests of *RUni2 (*.152 and .136*)*, and in the pre-test on *SUni2*. However, the overall data showed that only the pre-test data was normally distributed, suggesting the use of non-parametric data for the analysis as above. Reversing the claim about normality of data in the first research question, the cause for the non-normal distribution in these data could be due to the fact that there was clear skewness towards the minimum and there were two observations with a value of (0) in *SUni1*. When these two cases were excluded, no major change had occurred regarding normality. The detailed figures are shown in Table 4.7 below and the general observations are illustrated in Figure 4.2:

**Table 4.7 Tests of normality and reliability of CATSS (productive) SUni1 and RUni2 and both universities**

| School | CATSS test | Cronbach's Alpha | Shapiro-Wilk | | | N |
|---|---|---|---|---|---|---|
| | | | Skewness | Kurtosis | sig | |
| SUni1 | Pre | .750 | .718 | .205 | .025 | 51 |
| | Post | .849 | .279 | -.775 | .158 | |
| RUni2 | Pre | .569 | .285 | -.433 | .152 | 50 |
| | Post | .809 | .320 | -.084 | .136 | |
| 2Univs | Pre | .663 | .572 | .121 | .005 | 101 |
| | Post | .826 | .284 | -.499 | .050 | |

*$P > 0.05$ abnormal

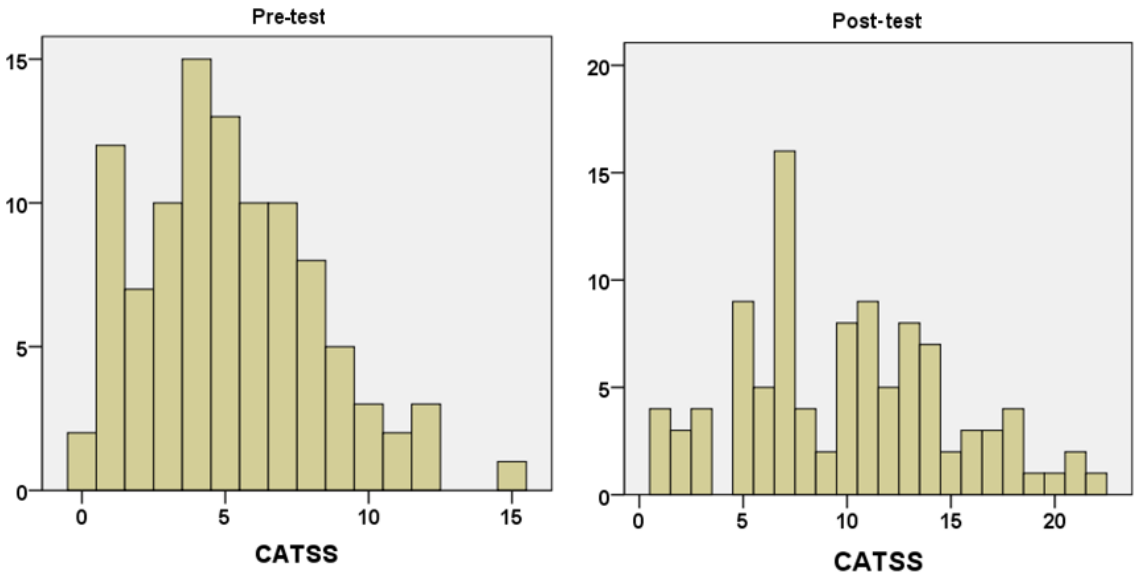



**Figure 4-3: Histograms of the AWL productive test CATSS**

### *4.5.2.2. Descriptive statistics of the productive test of the AWL:*

*The initial abilities of learners on presessional programmes:*

The scores on the productive knowledge test *CATSS* show that advanced EFL learners joining the presessional programmes lack the ability to produce the AWL. Learners from both universities have similar mean figures and their overall mean was 5.17/30. Even high achiever students failed to reach the maximum score and could only reach a score of 15 (50%). In real terms, this could mean that advanced EFL learners could only produce 98 words from the academic list, based on the formula stated earlier.

Table 4.8 below shows the pre-test means, the standard deviation, the minimum, maximum and range scores of the productive test CATSS.

**Table 4.8 Descriptive statistics of CATSS scores of SUni1 and RUni2, and both universities**

| | School | Mean /30 | ± SD | Minimum | Maximum | Range | *N* |
|---|---|---|---|---|---|---|---|
| CATSS pre scores | SUni1 | 5.137 | ± 3.44 | 0 | 15 | 15 | 51 |
| | RUni2 | 5.220 | ± 2.70 | 1 | 12 | 11 | 50 |
| | 2Univs | 5.178 | ± 3.08 | 0 | 15 | 15 | 101 |

*AWL productive improvement over the presessional programme:*

The scores of CATSS in Table 4.8 post-tests show higher improvement in the productive knowledge than the receptive test reported earlier. This is no surprise as learners had shown poor estimates in the pre-test, unlike the *receptive* or *recognition* test. However, it is unexpected that even high achievers failed to demonstrate the ability to produce two thirds of the academic list at the end of their academic programme. The mean score was actually below the mid score of the test. This could suggest that relying on academic exposure only on presessional programmes (i.e. with advanced learners) might not be sufficient for learning the entire academic list.

Learners at both universities showed similar results in improvement of productive knowledge of the AWL. The brief descriptive results are given in Table 4.9, and illustrated in Figure 4.4 below.

**Table 4.9 Basic descriptive statistics of means and standard deviation of both tests of CATSS**

| School | PreTest | ± SD | Post-test | ± SD | *N* |
|---|---|---|---|---|---|
| SUni1 | 5.137 | ± 3.44 | 9.706 | ± 5.17 | 51 |
| RUni2 | 5.220 | ± 2.70 | 10.180 | ± 4.89 | 50 |
| 2Univs | 5.178 | ± 3.08 | 9.941 | ± 5.01 | 101 |

Max = 30

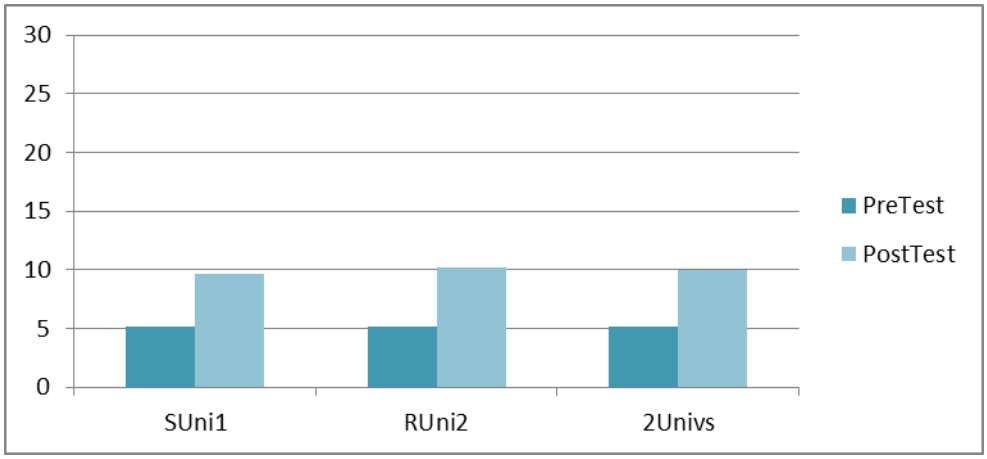



**Figure 4-4: Histograms of improvement of in the productive knowledge CATSS**

It is clearly noted from the figures that there was a significant increase in learners' scores since the beginning of the *presessionals*. The analysis of comparisons confirmed the significance among all learners. The analysis was performed by running the *paired sample t-Test* with the normal data (pre/post of RUni2) and *Wilcoxon test* with *non-normal* data (pre/post of SUni1). The comparisons show significant pre- and post-test differences at ($p<.001$). Table 4.10 below shows detailed figures of improvement levels in maximums, minimums, medians and difference;

**Table 4.10 Descriptive statistics of CATSS scores of SUni1 and RUni2, and both universities**

| School | CATSS | Mean +-SD 30/30 | Dif | P Value | *Min* | *Max* | *Median 30/* | *N* |
|---|---|---|---|---|---|---|---|---|
| SUni1 | Pre | 5.137± 3.44 | 4.57 | .000 | 0 | 15 | 5.00 | 51 |
| | Post | 9.706± 5.17 | | | 1 | 22 | 10.00 | |
| RUni2 | Pre | 5.220± 2.70 | 4.94 | .000 | 1 | 12 | 5.00 | 50 |
| | Post | 10.180± 4.89 | | | 1 | 21 | 10.00 | |
| 2Univs | Pre | 5.178± 3.08 | 4.77 | .000 | 0 | 15 | 5.00 | 101 |
| | Post | 9.941± 5.01 | | | 1 [a] | 22 | 10.00 | |

a. There were many observations with value (0) and (1), and therefore could not be excluded. A discussion about these unexpected low marks is in section

Although the significance may seem high, the results show modest levels of abilities to produce the AWL. This might be considered unexpected with advanced learners receiving comprehensive academic exposure during the presessional programmes. Besides, such a level of academic vocabulary capability might be insufficient for their graduate or postgraduate studies. Using the same formula to estimate the rate of increase, the words acquired during the learners' study was equal to 90.7 words. This represents a rate of 2.01 words a day over the course.

As stated earlier, exploring the relations between the productive knowledge in CATSS and the receptive knowledge in VLT will be addressed later. For the sake of organization

as explained before in section 4.5, the results of the *additional research question* about vocabulary size are discussed below.

### 4.5.3. Advanced learners' overall vocabulary size in the presessional courses

The relation between vocabulary size and knowledge is discussed later in 4.5.4. The results regarding this question are briefly reported because learners are not expected to increase their vocabulary size greatly in such a limited duration of time; i.e. the general vocabulary size of learners is a cumulative process that takes a long time.

*Reliability and normality*

The same procedure regarding reliability and normality check was performed on the data of XKlex test of vocabulary size. The test items were shown as reliable at more than .891 in the pre-test, and .779 in the post-test. Regarding the normality of distribution, the data from *RUni2* was shown as normal, but the data from *SUni1*, and the data as a whole, were confirmed as not normal. The detailed figures are shown in Table 4.11 below and the general figures are in Appendix F.

**Table 4.11 Tests of normality and reliability of XKlex (general vocabulary size) of SUni1 and RUni2 and both universities**

| School | XK-lex test | Cronbach's Alpha | Shapiro-Wilk | | | N |
|---|---|---|---|---|---|---|
| | | | Skewness | Kurtosis | sig | |
| SUni1 | Pre | .895 | -.154 | .106 | .554 | 51 |
| | Post | .713 | -.838 | .451 | .011 | |
| RUni2 | Pre | .876 | -.129 | -.682 | .558 | 50 |
| | Post | .868 | .163 | 1.163 | .527 | |
| 2Univs | Pre | .891 | -.409 | .180 | .168 | 101 |
| | Post | .779 | -1.134 | 2.037 | .000 | |

*Vocabulary size of advanced learners, and their improvement*

The scores of XKlex version (A) give estimations of a vocabulary size of about 4842 words among learners. The variation was ± 1254. The *RUni2* group showed higher vocabulary levels than the *SUni1* group. The figures are in table 4.12 below.

**Table 4.12 Descriptive statistics of scores of SUni1 and RUni2, and both universities**

| | School | Mean /10K | ± SD | minimum | maximum | *N* |
|---|---|---|---|---|---|---|
| XKlex pre scores | SUni1 | 4447 | 1340 | 700 | 7000 | 46 |
| | RUni2 | 5332 | 944.4 | 3200 | 7100 | 37 |
| | 2Univs | 4842 | 1254 | 700 | 7100 | 83 |

As stated earlier, the vocabulary size of learners is not expected to increase greatly during this limited time of presessional study. However, the results of the post-test could be used as confirmation of the estimate of vocabulary size reported by the pre-test test, especially since *XKlex* has been criticized for giving on overestimation of vocabulary size due to the way it works (see the previous discussion in section 3.4.1.3.). Nonetheless, the results give good indications about improvement since the current study targeted a large sample size.

The improvements and the significance of figures were revealed through *paired sample t-Test* (data from *RUni2*) and *Wilcoxon test* (data from *SUni1*) depending on the normality of data. The comparison analyses show significant pre-test/post-test differences at ($p<.002$).

Also, according to the XKlex test, students could learn at an average rate equal to more than 385 words during the programme. Table 4.13 below shows detailed figures of improvement levels in maximum, minimum, median and difference;

**Table 4.13 Descriptive statistics of scores of SUni1 and RUni2, and both universities**

| School | XKlex | Mean +-SD 10k | Dif | P Value | *Min* | *Max* | *Median 30/* | *N* |
|---|---|---|---|---|---|---|---|---|
| SUni1 | Pre | 4447±1340 | 385 | .036 | 700 | 7000 | 6300 | |
| | Post | 4832±1538 | | | 500 | 7000 | 6500 | 46 |
| RUni2 | Pre | 5332±944 | 418 | .003 | 3200 | 7100 | 3900 | |
| | Post | 5750±829 | | | 3600 | 8000 | 4400 | 37 |
| 2Univs | Pre | 4842±1254 | 389 | .002 | 700 | 7100 | 6500 | |
| | Post | 5231±1342 | | | 500 | 8000 | 7500 | 83 |

### 4.5.4. The relations between vocabulary size XKlex and the knowledge of the academic words in CATSS and VLT.

Scores from the three different tests (CATSS, VLT and XKlex) were used in order to answer the question addressed earlier about the relations between the abilities. Firstly, as mentioned previously in the procedures (section 3.4.4), the scores of the tests were transformed into percentages (/100%), since the tests have different maximum scores (please refer to data scoring and analysis procedures in Chapter 3). The descriptive statistics in Table 4.14 explain the variations between the tests, and are graphically illustrated in Figure 4.3 below.

**Table 4.14 Descriptive statistics of all scores of all participants transformed to 100%**

| | Test | original score | Mean | SD | *Min* | *Max* | *N* |
|---|---|---|---|---|---|---|---|
| Pre | CATSS | /30 | 17.26 | 10.26 | 0 | 50.00 | 202 |
| | VLT | /30 | 75.84 | 16.10 | 30.00 | 100.00 | |
| | XKles | /10k | 48.13 | 12.60 | 7.00 | 71.00 | 180 |
| Post | CATSS | /30 | 33.13 | 16.69 | 3.00 | 73.33 | 202 |
| | VLT | /30 | 81.84 | 15.27 | 36.00 | 100.00 | |
| | XKles | /10k | 53.32 | 13.28 | 5.00 | 80.00 | 184 |

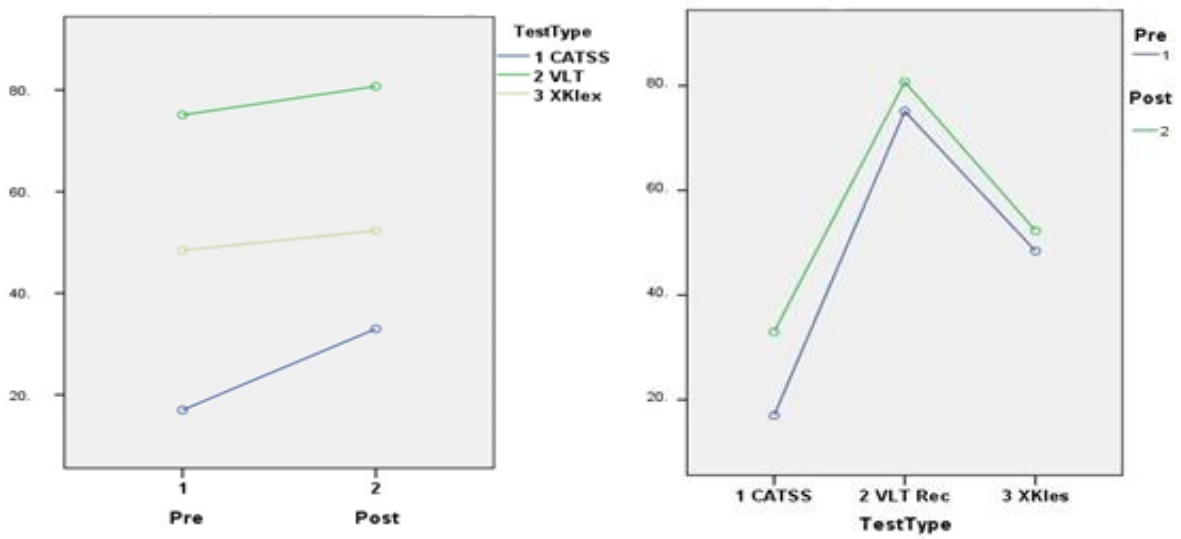



**Figure 4-5: Pre-test Post-test variations of all tests**

The figures in Table 4.14 show significant variation in the means of the three tests. The variation was obviously expected between the two measures that assess the academic words (CATSS and VLT) as they tend to assess different aspects of knowledge (simple recognition of the word vs. producing the word). As expected also, the two measures varied from the overall vocabulary size level (XKlex score) as it is supposed to target a greater vocabulary frequency.

Statistically speaking, the data was analysed by the two-way repeated measure *ANOVA* with Greenhouse-Geisser correction. The analysis was performed on the ranked data to avoid violations regarding normality (Beasley & Zumbo, 2009; Field, 2013) and showed significant difference between the different tests F (1.398, 322) = 101.292, $P<0.01$.

In terms of improvement, the data showed a significant difference between the pre-tests and post-tests F (1, 165) = 98.611, $P<0.01$.

The graphs and the figures of the analysis show a significant interaction between the pre-tests and post tests of CATSS and VLT, and CATSS and XKlex ($p<0.001$). The post-test scores compared to pre-test scores were significantly higher for CATSS (mean difference between pre and post=15.87) than those of both VLT (mean difference = 6.00) and XKlex (mean difference =5.19). No interaction was found between the pre-tests and post tests of VLT and XKlex ($p>0.05$). See the brief summary of means and difference of increase in Table 4.15, and interactions in 4.16 below:

**Table 4.15 Descriptive statistics of scores of all measures**

| | Pre | Post | Mean Difference | *Z* | Sig. (2-tailed) |
|---|---|---|---|---|---|
| CATSS | 17.26 | 33.13 | 15.87 | -11.324 | .000 |
| VLT | 75.84 | 81.84 | 6 | -7.116 | .000 |
| XKles | 48.13 | 53.32 | 5.19 | -4.337 | .000 |

Pairwise comparisons based on Wilcoxon Signed Ranks Test

**Table 4.16 Interactions between measures**

| | Sig |
|---|---|
| CATSS ↔ VLT: Pre*Post | .000 |
| CATSS ↔ XKlex: Pre*Post | .000 |
| VLT ↔ XKlex: Pre*Post | .132 |

*Correlation between different abilities:s*

Generally, the three tests correlate significantly with each other, in pre-tests and post-tests. The figures also show a higher correlation concerning abilities in the post-test than the pre-tests. The correlations show a significance level at ($p>001$). The Tables below summarize the correlations:

**Table 4.17 correlation figures of pre test**

| | | N | CATSS | VLT | XKlex |
|---|---|---|---|---|---|
| CATSS | Correlation Coefficient | | 1.000 | .536**** | .281**** |
| | Sig. (2-tailed) | 202 | | .000 | .000 |
| VLT | Correlation Coefficient | | .536**** | 1.000 | .479***** |
| | Sig. (2-tailed) | 202 | .000 | | .000 |
| XKlex | Correlation Coefficient | | .281**** | .479**** | 1.000 |
| | Sig. (2-tailed) | 184 | .000 | .000 | |

**Table 4.18 correlation figures of post test**

| | | n | CATSS | VLT | XKlex |
|---|---|---|---|---|---|
| CATSS | Correlation Coefficient | | 1.000 | .612**** | .422**** |
| | Sig. (2-tailed) | 202 | | .000 | .000 |
| VLT | Correlation Coefficient | | .612**** | 1.000 | .472**** |
| | Sig. (2-tailed) | 202 | .000 | | .000 |
| XKlex | Correlation Coefficient | | .422**** | .472**** | 1.000 |
| | Sig. (2-tailed) | 180 | .000 | .000 | |

**. Correlation is significant at the 0.01 level (2-tailed).

Both learners' mastery abilities of the AWL (receptive and productive) were found to correlate significantly in the pre-test (.536). An even better correlation is found between performances of two of the post-tests (.612). Some correlation was found between the knowledge of the academic words (productively and receptively) with overall vocabulary size. The correlation between CATSS and XKlex was higher in the post-test (.422) than in the pre-test (.281). The correlation was almost the same between XKlex and VLT in the pre-test and post-test (.479 and 472).

The higher levels of correlation between CATSS and VLT are expected because the tests tend to measure two aspects within the AWL knowledge continuum (recognition and production). The correlations between overall vocabulary knowledge and the two performances of the AWL are also predictable because the knowledge of the AWL is contained within the 10,000 most common vocabulary frequencies. However, as seen, the

correlation was lower between the vocabulary sizes of learners, and their productive and receptive knowledge of the AWL.

### 4.5.5. Advanced learners' free productive ability of the AWL: the lexical richness in free writing.

As in the previous questions, this research question is composed of three components; a) the initial AWL lexical frequency in learners' writing at the beginning of the course, b) the level of the AWL lexical coverage in their writing at the end of the programme, and c) the relation between their knowledge and ability to use the AWL and their lexical choice in free writing.

As explained in the procedures, the written data of a number of students was collected twice: at the beginning and just before the end of the course. As mentioned, the data of a sample of 8 students from one class was used for this analysis of the AWL frequency in their writing. The pre- and post-writing passages were typed into *MS Word*, and then were run through the *vocabprofiler*. As stated in the procedures, misspelled or unclear words were shown to a native speaker validator to be corrected into possible proper words to be recognized by the *vocabprofiler*.

Only 10 words (5 words from one participant) out of the total writing of all learners in the pre- and post- writing needed correction. The entire set of words corrected is in Appendix I.

#### *4.5.5.1. Descriptive reports from the vocabprofiler on learner' writing and initial AWL frequencies at the beginning of the presessionals:*

Since the data of only 8 students was found suitable for the analysis, it was believed that presenting all the data would be more meaningful than showing general statistical representations. Below is a qualitative report from the *vocabprofiler* of all participants

which summarizes the vocabulary frequencies in learners' writing. Only the data relevant to the study (AWL figures) was included in the table, as stated previously in procedures. See Table 4.19 below:

**Table 4.19 Descriptive reports from the vocabprofiler (pre-test) based on equal word count of writing produced by all participants:**

Definitions of families, tokens, and types are found in 2.4.4.

| Participants | Word Count | 1k+2k % in Text | AWL | | | |
|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percentage |
| 103 | 200 | 93.00% | 6 | 6 | 9 | 4.50% |
| 105 | 200 | 95.02% | 5 | 5 | 5 | 2.49% |
| 113 | 200 | 93.56% | 4 | 4 | 4 | 1.98% |
| 114 | 200 | 93.50% | 11 | 11 | 13 | 6.50% |
| 122 | 200 | 91.01% | 6 | 6 | 9 | 4.46% |
| 129 | 196 | 85.86% | 12 | 12 | 14 | 7.07% |
| 138 | 200 | 84.50% | 16 | 21 | 24 | 12.00% |
| 144 | 200 | 92.50% | 3 | 3 | 5 | 2.50% |

**Table 4.20 Descriptive reports from the vocabprofiler (pre-test) including the entire writing passages of all participants**

| Participants | Word Count | 1k+2k % in Text | AWL | | | |
|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percent/All Words |
| 103 | 307 | 92.83% | 10 | 10 | 13 | 4.23% |
| 105 | 226 | 94.74% | 7 | 7 | 7 | 3.07% |
| 113 | 227 | 94.43% | 4 | 4 | 4 | 1.72% |
| 114 | 235 | 93.99% | 12 | 12 | 14 | 6.01% |
| 122 | 203 | 91.22% | 9 | 9 | 9 | 4.39% |
| 129 | 196 | 85.86% | 12 | 12 | 14 | 7.07% |
| 138 | 288 | 86.15% | 16 | 22 | 30 | 10.38% |
| 144 | 281 | 91.46% | 9 | 9 | 12 | 4.27% |

The report gives figures about percentages of GSL 2000 words in the written texts, and then counts how many academic word families, types and tokens appeared in the texts, and what percentage was from the AWL. As stated in the procedures, two analyses were run; the first included only the first 200 words of each written passage, and the second included the whole length of the compositions produced by the learners.

It might be assumed that the second analysis (measuring the frequency of all words produced) might not produce consistent results since participants were unlikely to produce essays of similar word length, and this might affect the frequencies. However, the second analysis was performed because there was a clear similarity in the patterns of frequencies between the analysis of the 200 words and the overall length of the essays. Also, some of the data which emerged was found to have some possible implications that will be discussed shortly.

The study of Morris and Cobb (2004) mentioned earlier in the thesis argues that the ideal percentage of the AWL in free academic writing is above 5%. Of course, with the significant correlation between vocabulary size and writing that has been found in literature, higher levels of vocabulary do not necessarily guarantee better writers, as Stæhr (2008) previously commented. Bearing this in mind, Table 4.20 above shows that five of the eight students had AWL percentages lower than 5% in free writing. The two lower achiever writers had only 4-5 academic words out of about 227 words essay. This could mean that some advanced learners in their last stage of English study struggle to use academic words freely in writing. The *word types* (create and creative are two *types* of one *family* word) range between 4 to 21 with a median of 9.5, and the families between 4 and 16 with a median of 10.6 words. The entire AWL *types* that each writer had mentioned in his essay are available in Appendix G.

*4.5.5.2. Descriptive reports from the vocabprofiler on learners' writing and the AWL frequencies at the end of the presessionals:*

Generally, apart from participant 138, learners have either produced more, or used the same range of, academic words in their writing in the second task. This is based on the

count of the word *families* written in their compositions. There was some decrease in AWL use in the second writing with two participants (138,114); however, based on the analysis of the entire essay, they had used more AWL families. Although it is not possible to generalize from a limited sample such as this, the only pattern that could be found is that the five participants who had lower AWLs than 5% in the first writing had increased this count in the second writing. Table 4.21 below shows a detailed descriptive report from the *vocabprofiler* of the second writing, and Figure 4.6 illustrates the that the comparisons in a histogram.

**Table 4.21 Descriptive reports from the vocabprofiler (post-test) based on 200 words of the written compositions**

| Participants | Word Count | 1k+2k % in Text | AWL | | | |
|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percent |
| 103 | 200 | 92.58% | 9 | 9 | 11 | 5.45% |
| 105 | 200 | 93.50% | 5 | 6 | 7 | 3.50% |
| 113 | 200 | 93.07% | 4 | 5 | 5 | 2.48% |
| 114 | 200 | 92.04% | 7 | 7 | 9 | 4.48% |
| 122 | 200 | 86.70% | 13 | 13 | 14 | 6.90% |
| 129 | 200 | 88.12% | 12 | 12 | 13 | 6.44% |
| 138 | 200 | 89.50% | 7 | 9 | 9 | 4.50% |
| 144 | 200 | 91.00% | 7 | 8 | 9 | 4.50% |

**Table 4.22 Descriptive reports from the vocabprofiler (post-test) based on the entire compositions produced by participants**

| Participants | Word Count | 1k+2k % in Text | AWL | | | |
|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percent/200Wds |
| 103 | 592 | 91.43% | 24 | 27 | 37 | 6.22% |
| 105 | 453 | 93.37% | 12 | 14 | 15 | 3.31% |
| 113 | 491 | 94.73% | 12 | 13 | 14 | 2.81% |
| 114 | 354 | 91.89% | 12 | 12 | 16 | 4.64% |
| 122 | 380 | 84.65% | 32 | 33 | 36 | 9.21% |
| 129 | 356 | 87.43% | 24 | 24 | 27 | 7.54% |
| 138 | 517 | 89.57% | 22 | 25 | 29 | 5.60% |
| 144 | 491 | 89.53% | 20 | 21 | 25 | 5.13% |

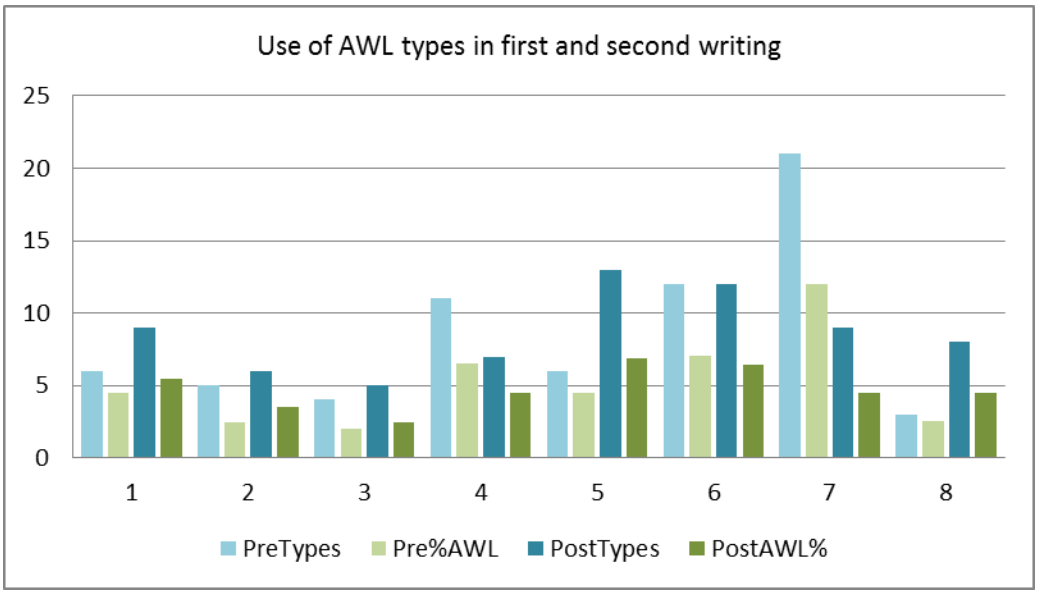



**Figure 4-6: Use of AWL types in first and second writing**

Although there was a clear increase in the use of the AWL in the writing of the second occasion as reported, only three writers exceeded the 5% figure suggested earlier. The same lower achiever writers hardly produced 4-5 academic words in their second writing even though it was a longer essay. The AWL word *types* and *families* in the second writing ranged from between 5 to 13 with a median of 8.5 in word *types* and 7 in the word *families.*

Of course, as stated, from a limited sample such as this, not too many reliable conclusions can be drawn. However, the main conclusion that could be reached is that some advanced learners at the end of their presessional course, just before continuing their academic study, may have a poor AWL lexical profile in their academic writing; i.e. with respect to the ideal percentage of the AWL in written texts suggested earlier by Morris and Cobb, the above report of the second writing shows that 5 participants out of 8 failed to reach the 5%. However, as seen above, 3 writers have a percentage that was near the ideal figure (i.e. the 5%) and reached around 4.50%. As argued before, writing and vocabulary

size correlate significantly, but, as Stæhr commented, learners with higher lexical levels do not necessarily make better writers.

*4.5.5.3. The relations between AWL abilities and the lexical choice in free writing:*

Of course, with such a small sample size, the data would be insufficient to perform reliable statistical analysis and draw conclusions due to the influence of the population, outliers etc. The data of a small number of participants simply cannot be generalized confidently. For this reason, further details about all participants were reported earlier, and analyses were made. However, since this study performed three multiple measures as well as collecting two writing samples from learners, it was thought that the data regarding the relationship between AWL frequencies and the productive and receptive tests as well as the vocabulary size was still worth investigating, at least for building some grounds for the second study and for future studies.

Generally speaking, as expected, the analysis revealed that there was no clear correlation or interaction between the three tests and the AWL profile in academic writing with the learners.

Table 4.23 below gives descriptive statistics for the eight students in all previous measures: CATSS, VLT (or Rec) and XKlex, based on the 200 word count analysis.

**Table 4.23 Descriptive reports of learners test scores and their AWL lexical frequencies:**

| Students | VLT or REC | | XKLEX | | CATSS | | AWL | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | pre | post | Pre | Post | Pre | post | Family | | Percentage | |
| 103 | 73.3 | 90 | 51 | 58 | 10 | 40 | 6 | 9 | 4.5 | 5.45 |
| 105 | 83.3 | 96.6 | 50 | 59 | 20 | 56.6 | 5 | 5 | 2.49 | 3.5 |
| 113 | 83.3 | 86.6 | 66 | 66 | 6.6 | 40 | 4 | 4 | 1.98 | 2.48 |
| 114 | 30 | 73.3 | 56 | 52 | 6.6 | 23.3 | 11 | 7 | 6.5 | 4.48 |
| 122 | 80 | 90 | 40 | 51 | 23.3 | 16.6 | 6 | 13 | 4.46 | 6.9 |
| 129 | 90 | 90 | 57 | 57 | 10 | 36.3 | 12 | 12 | 7.07 | 6.44 |
| 138 | 96.6 | 100 | 66 | 51 | 23.3 | 23.3 | 16 | 7 | 12 | 4.5 |
| 144 | 73.3 | 90 | 44 | 67 | 21 | 26.6 | 3 | 7 | 2.5 | 4.5 |

It is impossible to identify any pattern or sign of interaction between the scores of the different tests and the use of the AWL freely in writing, based on such a small sample size. However, as mentioned, since the data was thought to be valuable to perform further investigations to look for any signs of possible interactions, a follow-up analysis was performed and merged all the data from the *pre* and *post* writings as well as the pre- and post-tests of the other tests (n = 16); i.e. the data of the four measurers (CATSS, VLT, XKlex and AWL frequency) of the second occasion (post-test) was treated as extra data of 8 students.

The results showed positive correlations between the count of AWL *families* and the AWL receptive test VLT (Kendall's tau =.434, $p<.05$), and between the AWL productive test CATSS (Kendall's tau =.368, $p<.05$). The analysis was based on all word counts of the essays that learners produced. The items were checked for reliability using *Cronbach's Alpha* test of reliability and the figure was 0.59. This reliability figure is expected based on this small sample.

Although the results cannot be generalized to any great extent, they at least provide an indication about an expected correlation between the receptive and productive AWL abilities with the free use of academic words in writing.

Also, as mentioned before, these results support the findings of Laufer and Nation (1995), who used similar vocabulary tests and investigated more frequency range of words (the second 1,000 words, the third 1,000, the fifth 1,000, the University Word List, and the tenth 1,000). At the time of their study, the test was based on the University Word List (Xue and Nation, 1984). Meara (2005) conducted a follow-up study and raised some concerns about the generalization of Laufer and Nation's study regarding the correlation

between the lexical frequency profile LFP with an independent measure of vocabulary knowledge. His remark was that very large groups comprising students with a wide range of vocabulary sizes should be investigated to reach a confident generalization. One of his logical arguments concerned the uncertainty of the direct relation between passive and active vocabulary measurement. Laufer (2005b) responded to the criticism stating that LFP does not tell us whether learners can produce certain words when prompted to do so, but rather tells what proportion of frequent vs. infrequent vocabulary they choose to use in their writing.

The position of this thesis study regarding the above research timeline is that the study found a possible correlation in a specific frequency range (AWL) between productive and receptive tests with the lexical frequency of academic words in writing, with a specific level of learners (advanced EFLs).

Lastly, there was concern about the fact that there might be re-occurrences of the same academic words in both writings. This may affect the analysis in the sense that some learners might have better scores because of repeated words, while in fact it does not give a clear indication of improvement. All word types that were produced in both pieces of writing of all learners were examined with this in mind. Based on the 200 word count, the recurrences varied between 0 and 2 words, and the variety was in line with the count of AWL items produced. This should not have an effect on the outcomes of the analysis. The AWL items produced by each writer are available in Appendix G.

## 4.6. Summary of the chapter

This study investigated the levels of vocabulary size and abilities regarding the academic words by advanced EFL learners at their last stage of language study, and their degree of improvement in these abilities that the presessional programmes allow.

Due to the variations in the results previously reported in the literature about advanced students, plus the uncertainty regarding their capabilities, and the incremental nature of vocabulary knowledge, this study used multiple measures to assess the abilities of learners. As well as this, the lexical frequency profile LFP of the academic words in writing was checked twice: at the beginning of the presessional programme and at the end, in order to determine the ability of advanced students to use the AWL freely in writing, and relate their LFP to their levels of vocabulary knowledge. Similar students from two universities were included in the study in order to ensure higher levels of confidence and stability in reporting the results and making generalizations.

The main results show that advanced learners on presessional programmes show higher levels of *receptive* knowledge of the AWL, mastering at least two thirds of the list. There were some lower achieving learners who were able to recognize less than one-third of the academic words even though they were at this relatively advanced level of proficiency and were about to study for their degree. The overall improvement during the programme was significant at a range of 34.2 words. The productive ability was generally low. Learners could only produce about 20% of the academic list, while the higher achievers barely produced half of the list. There were some lower achievers who failed to produce any academic words. The improvement (increase) of the knowledge of the AWL during the programme was significant; however, even high achievers failed to produce two

thirds of the list after the course. This could suggest that relying on academic exposure only on presessional programmes (i.e. with advanced learners) might not be sufficient to cover the entire academic list. The general vocabulary size of learners was around 4842 words, which increased to 5231 words by the end of the programme. The tests correlate significantly with each other, in both pre-tests and post-tests, and in the change difference (increase rate). Productive knowledge (CATSS) was also found to interact significantly with the receptive test (VLT) and vocabulary size (XKlex). No interaction was apparent between performance on the VLT and the XKlex. Lastly, some learners were found to have poor abilities regarding producing the AWL in free writing. The improvement in the use of the academic words in the second writing task at the end of the programme is noticeable, and the correlation between the tests of vocabulary and the lexical choice of academic words in free writing was found to be significant with some limitations regarding sample size.

The results of this study are discussed further in Chapter 7.

Lastly, it was mentioned in the methodology chapter that XKLex test (Masrai, 2009) was used in *study one* and *study two* to check learners' general vocabulary size. It is useful to note that at the time of the study discussed in this chapter, it was believed that the pattern in the construction of the items of the test (the order of the *non-words* in the test sheet) was not considered an issue, since the test was trialled and validated by (Masrai, 2009) and produce consistent results. Although Masrai's study involved Saudi participants, he noted that AlHazemi's (1993 *cf* Masrai, 2009) found that Saudi learners use a high rate of guesswork in answering unknown words. The pattern used by Masrai, I believe, was to make the test easier for marking by the examiners. At any rate, it was decided in the

second study to alter the order of the *non-words* in the test and make no pattern in its construction, since all participants of *study two* are all Saudi learners, considering AlHazemi's findings. This is discussed further in the following chapter (see section 5.2.3).

# 5. Chapter Five

## The Second Study:

## Investigating the Learning of the Academic Word List by Learners experiencing Two Different Methods of Teaching on an Advanced Intensive English Programme.

## 5.1. Introduction.

This study explores the learning outcomes resulting from two methods of learning the academic words on an advanced intensive English language programme. It has been discussed on many occasions earlier in the thesis that advanced students often learn vocabulary incidentally due to their capability and level of proficiency. Many studies reviewed in this paper have discussed the issues of incidental learning of vocabulary vs. explicit or enhanced teaching of words, either with beginners and/or advanced EFL learners. Some arguments have been put forward related to studies involving advanced students regarding effective methods of teaching vocabulary, and studies of incidental or explicit teaching (see sections 2.3.2.1 and 2.3.2.2). For example, limitations were discussed concerning the targeted words used for the studies (Elgort, 2011; Folse, 2010; Joe, 2010; Sonbul & Schmitt, 2009; Webb, 2009), the measures used (Alemi & Tayebi, 2011; Webb, 2009), and/or the concerns about research design and the ability to generalize results (Alemi & Tayebi, 2011; Sonbul & Schmitt, 2009). In order to minimize the negative impact of these issues, this study applied multiple measures in order to assess the learning of academic words by learners studying in two different groups, and receiving two different methods of vocabulary teaching. Some students had normal exposure to academic words through the academic skills (mainly listening and reading), while others received an extra outside-class enhancement of the academic words, as discussed further in the methodology. Keeping in mind the issue of the complexity of vocabulary knowledge, this study therefore attempts to compare the results of similar students experiencing two approaches of teaching using more in-depth measures to explore the learning outcome to which a certain approach contributes.

With respect to the discussion above, this study aims to address the following research question:

> **RQ 2-** Which approach to vocabulary teaching (incidental/explicit) is likely to lead to the more significant gain?

As discussed in section 3.3, an additional sub-research query has evolved, and a number of research hypotheses have been put forward. In order to respond to the research questions and research hypotheses, this study attempts to:

1. Checked the vocabulary size of learners before their entry into the academic level (see information about participants below),
2. Assessed their knowledge of the AWL; productively and receptively, and the lexical richness of the AWL in their writing.
3. Since some learners in this study learn academic words explicitly, the study attempts to determine how explicit teaching would contribute to learners' knowledge of the academic words (receptive and productive), and how would it affect their lexical choice in writing.

## 5.2. Method

### 5.2.1. Participants

The participants of the main study were EFL Saudi male learners studying at the Preparatory (or foundation) year (PY) at King Saud University (KSU) in Riyadh, Saudi Arabia. The PY could be described as the largest EFL PY programme in the world as it accommodates up to 8000 students, both male and female as well as 1500 students from a neighbouring community college. The PY aims to develop the learners' language skills (the receptive and the productive skills) and prepare them for communication at the

academic level. The PY administers a proficiency placement test before the term begins, and assigns learners to their groups depending on their language proficiency. There are *six* levels that start from level one (beginners) to six (high achievers). In each level there are a number of classes and student groups, with no fixed number of classes at each level. The targeted sample in this study was male students from the groups in the *fifth* level. This particular level was chosen after a number of considerations. Beginner levels were thought to be of less interest in terms of academic vocabulary, as their proficiency levels were basic (see below). Although the fifth and sixth levels receive the same input in the PY curriculum (as seen below), the sixth level was excluded from the experiment as, at the time of the study, it had only two classes as usually in Term 1 of the year, there are fewer advanced students. It was believed that such a sample size of students was very low for this kind of experiment. In other words, since the study planned to use many tests with the same students over a period of time, it was thought that a small sample of students would be affected if some students had not shown up in all of the tests on both occasions (i.e. pre- and post-tests). Students at the sixth level were very advanced anyway, and it was believed that they knew much of the AWL.

The learners receive 20 hours of English classes every week in two terms (600 hours in total). Each term is divided into two halves (4 quarters a year). Each term students are tested once at the end of the term, and also at mid-term.

Students receive English learning 'input' in the PY programme according to their levels of proficiency. The content of their assessments also depends on this. To explain further, the beginning and the intermediate levels receive general English classes (each quarter finishing one level within), and the advanced levels receive academic education (input).

Better achievers at any level can take IELTS or any other recognized exam and move to the level that suits them, or even register for fewer English hours if they reach the required mark for their schools. Levels *five* and *six* are equal to IELTS 5.0 according to the administration of the PY using their local test, with level *six* being slightly more advanced and mostly from medicine.

87 students from two learning groups (8 classes in total) in level five participated in the study. Each of the two groups experienced a different method of teaching academic vocabulary (see Section 2.3.2.2.). There were 44 students in treatment groups (henceforth: direct AWL or DG) and 43 in the control group (henceforth: non direct or incidental AWL or NnDG).

Table 5.1 below provides more detailed information about the participants, based on the questionnaire sheets.

**Table 5.1 Overview about the learners participating in this study:**

| | | Major | | | | PG, or UnderG | |
|---|---|---|---|---|---|---|---|
| | | Humanities | Medicine | Science | N/A | PG | UG |
| DG | | 33 | 8 | - | 3 | 44 | - |
| NnDG | | 27 | 11 | 1 | 4 | 43 | - |
| Total | 87 | 60 | 19 | 1 | 7 | 87 | 0 |

It might be assumed that the major of study might have some influence regarding motivation and, therefore, AWL learning, i.e. medical students might be considered more motivated than learners from other majors due to the difficulty of their subjects. The inequality between the two groups regarding majors might cause concern regarding data analysis. However, the two groups, as seen above in Table 5.1, have a fair distribution among participants in terms of major of study.

Finally, several more students in level five actually agreed to participate in the study; however, only the above number qualified to be included in the study because not all students took all of these multiple tests.

### 5.2.2. Procedure, materials used, and a discussion of some ethical issues.

As discussed earlier, advanced students at the PY were taught academic English in their 20-hour weekly programme. According to the Curriculum-Assessment Framework (CAF) set by the PY, levels *five* and *six* receive EAP and ESP classes, as well as classes on study skills and IELTS preparation. It was mentioned that one of the main aims of the study is to compare the learning outcomes of applying *explicit* learning of academic vocabulary to advanced learners and *incidental* learning of the AWL from academic exposure. The incidental learning groups (NnDG) received normal input according to the CAF plan set by the PY. The *explicit* or direct AWL groups (DG) received extra treatment, which consisted of outside classroom activities that enhance the academic words.

The procedure and planning for this extra activity was as follows:

First, it was believed that targeting only a small number of words from the AWL and choosing them for teaching and then assessment might not present results that could be relied on and confidently generalized. In other words, targeting fewer words in a large scale study involving a large number of students could result in difficulty in detecting any variations which might result from two methods of teaching. Besides, the study used general assessment of the AWL, and therefore, the targeted range of words for teaching needed to be broad to ensure that what was assessed was covered or taught. Thus, 240 words of the AWL (out of 570) were chosen for the treatment. The 240 words are the most common academic words within Coxhead's (2000) academic word list. The words

were divided into groups of 20 and presented to the students during 8 weeks of the 12-week programme. With the help of teachers, the twenty words were presented in exercises which drill the word meaning, word families and collocations, and help develop learners' abilities to produce natural use of words in writing. These exercises were done outside the class as homework or as independent study. Although the same list was presented to all of the groups in DG, the teachers were advised to follow exercises, tasks and techniques of a certain textbook that focuses on the AWL to ensure as similar an input as possible for all learners. The textbook was *Focus on Vocabulary: Mastering the Academic Word List* (Schmitt & Schmitt, 2005). The types of tasks and exercises of this textbook, which teachers were advised to follow, vary between multiple choice (choosing the targeted words from a number of definitions, correct uses or synonyms), matching synonyms or accurate collocations, filling in gaps with missing words or affixes, and choosing the right form of words from their base or most common shape. It should be noted that the construction of the tasks for inclusion in the treatment was not the same as the construction of the research test tools.

Learners were asked to show their work to their teachers. These classroom practices were supervised by the researcher and both teachers and students filled in a survey which is discussed here in a different chapter. The role of the survey in this study was to monitor any unexpected results or outliers in specific individuals. Having two similar participants in one school and one level, some of whom had the privilege of extra input, may raise some ethical issues. In defence of this, the following points need to be made:

1- At the first meeting with students before the study began, all learners, were aware that participating in this study was completely voluntary and that they were free to withdraw from the project at any time.
2- They were aware that they could move to any group within level five (i.e. choose between DG or NnDG groups). This action was approved by the PY administration.
3- At the PY, teachers are passive regarding the curriculum of the PY, and mostly follow a similar syllabus.
4- As stated earlier, the activities of the treatment (learning the AWL) were given outside the classroom as homework, and did not affect the instruction of classes at the PY.
5- At the PY, students have web forums and social networks where they can share ideas and thoughts about the subjects they learn on the programme. These forums are popular amongst students, and learners of the control group knew about the project and could have asked to join for extra input of the AWL.
6- Higher education is free in Saudi Arabia, and learners, therefore, are less concerned about extra learning activities that are given to only some of them. i.e. if learners were paying for their studies, they may be more concerned with differences in classroom practices.

Teachers were asked to collect any exercises presented by their students, so that they would stay motivated. At the PY, as CAF states, 20% of the mark is given by the teacher for active and hardworking students. See Table 5.2. below for the breakdown of scores per semester set by the KSU, PY; the IEP Curriculum-Assessment Framework.

**Table 5.2 The IEP Curriculum-Assessment Framework CAF set by the PY**

| Evaluation | Proportion of marks |
|---|---|
| **Midterm exam** | 30% |
| **Self Learning** | 10% |
| **Continuous Assessment** | 10% |
| **Final Examination** | 50% |
| **TOTAL** | **100%** |

All students from groups DG and NnDG took the multiple assessments of this study (the pre-test series) on week two of the Autumn-term 2010, and in week 10, they had the second series of post-tests. It should be noted that there was a holiday gap of two weeks after the first midterm.

### 5.2.3. The instruments.

The instruments used for this study were the same multiple instruments used for the preliminary study (the first section in this chapter). As previously mentioned, on a number of occasions in this thesis (see e.g. 2.4.2), multiple measures of words are crucial because of the nature of vocabulary mastery. That is one of the main aims which the first study in this chapter addresses. However, in this study, the multiple measures assessment is used to investigate the learning outcomes of similar learners who experienced different methods of instruction. The use of multi-testing of all students is an appropriate means for detecting any variations that may occur in students studying in two different learning groups.

There were no modifications (except the construction of the post-test of XK-lex as discussed below) to the instruments, and the same procedure was followed as in *the first study* (Chapter 4). There was no need to translate the instructions of the tests into Arabic

(the first language of all participants) as the students were at an advanced level. The multiple measures briefly are; the XK-lex (Meara & Milton, 2006), the vocabulary level test VLT (Schmitt, Schmitt, & Clapham (2001), CATSS or the Computer Adaptive Test of Size, a paper and pen version, (Batia Laufer, personal communication) and the vocabulary profiler (*Vocabprofiler;* Cobb, 2008). The only amendment was losing the hidden pattern (order of *non-words*) that the construction of XK-lex originally has, as discussed at the end of Chapter 4. It was decided at the time of the study to avoid any complications that might occur due to having a hidden pattern in the construction of the test that might influence learners' guessing. Thus, the *post-test* (version B) of XK-lex test had no hidden pattern in its construction.

In Table 5.3 below is a summary of the instruments used and what they were intended to assess, and the timing of tests. The table is similar to Table 4.2, but includes information from the study described in this chapter;

**Table 5.3 A summary of the tests and procedure :**

| Instrument | What intends to test | Participants | Time |
|---|---|---|---|
| Version **A** of the VLT | Receptive knowledge of the AWL. | All students | Week **2** of the course |
| Version **B** of the VLT | Receptive knowledge of the AWL. | All students | Week **10** of the course |
| Adopted version of *CATSS* | productive knowledge of the AWL | All students | Weeks 2 and **10** of the course |
| Writing Task **A** and vocabprofiler | Lexical richness of the AWL in writing | 6-8 students | Mid Term |
| Writing Task **B** and vocabprofiler | Lexical richness of the AWL in writing | 6-8 students | Final |

The writing tasks procedure in this study, however, was different from the first study. Due to logistics, it was not possible in this study to set writing tasks for participants and collect them during study time. However, students and PY administration agreed for a sample of the writing data in the mid-term and final assessment to be used in the study.

The writing tasks were supposedly academic writing questions which discussed similar topics. However, at the time of collecting the post-writing, some learners had a task that did not closely match the others. Below are the writing tasks:

**Task 1:** Write at least 150 words on the following topic:
" Write an essay explaining how you would solve the environmental problems in your city" Give at least two examples of problems and possible solutions".

**Task 2 A**: Write at least 200 words on the following topic:
"New inventions, such as the internet, mobile telephones and the Global positioning system have changed the world we live in. Discuss the advantages and disadvantages of new technology. Give examples to support your answer".
**Task 2 B:** Write at least two to three paragraphs on the following topic:
"Write two or three paragraphs about your favorite place.
Say: 'where this place is, what the place looks like, and why it is your favorite'."

During the research design of the study, it was assumed that the writing tasks would be mostly similar and therefore suitable for comparison. However, at the time of the data collection, it was hard to find sufficient written data that represented the two learning groups which could to be used for comparison. Since the targeted level of students were advanced learners, many of them passed IELTS exams later in the term and therefore were not required to do the writing test according to the PY policy. Apart from this, not all learners had participated in the other measures of the study, neither had they had done both the pre- and post-writing. Furthermore, some writing data had to be excluded due to difference in task genre. i.e. not all writing tasks were academic since learners within level five were tested in the final test in classes according to their study major. Some learners from level five took an advanced final test, and other learners took a test that was less advanced in order to match learners from level 4. Thus, finally, the only written data that was suitable for comparison between the two learning groups, and was available at

the time of the study, was from learners who took Task 1 and Task 2 *B*. The implication of this is further discussed in the results.

## 5.3. Scoring and coding procedures and data Analysis

This study follows the same scoring and coding procedures of the *First Study* in Chapter 4, and is explained in the Methodology chapter.

Generally, the responses of the three measures were itemized and transformed into numerical codes (1 for correct and 0 for missing or incorrect responses), then entered into *Excel* in order for easy processing with SPSS statistical package. The final scores consist of the sum of all the correct answers for each measure.

Each test has a particular scoring procedure as explained in section 3.4.3.

Regarding the data analysis, the five process stages mentioned in the Methodology chapter and used in the *First Study* were repeated in this study, i.e. the *reliability* and normality tests of measures, the transformation of scores into percentage, the *two-way ANOVA* to compare scores of the pre-tests then the post-tests of each measure with respect to group effect, and then the *t-test* of each measure pre- and post to retrieve the means, and *the Correlation Coefficient* to examine the relation between the means of these were performed.

As also mentioned in Chapter 3, since this study involved participants from two groups experiencing two methods of learning the AWL (the DG and NnDG groups), a sixth process of data analysis was performed. A two-way mixed design *MANOVA* was run to compare all scores in the two learning conditions to determine if the averages were statistically different. The *t-test* (or any of equivalent) was performed to compare the means of the two groups in each test type, as illustrated in Figure 3-2 presented earlier.

As mentioned before, these procedures aim to demonstrate:

- The overall score results of all tests, means, minimum and maximum figures as well as standard deviation
- The correlation between the measures, as well as the effect of scores on learners' use of the AWL in free writing
- The significance of an increase or decrease, and the degree of development of vocabulary knowledge in both groups.
- The variations in the improvements resulting from the two different methods of exposure

### 5.4. Results and discussion

The main research question and hypotheses involve exploring:

- Learners' improvement regarding the receptive or recognition level of the AWL, as well as the ability to produce the AWL, based on the learning group they are involved in.

As discussed in section 3.3, the main research question of this study suggested a further additional sub-research query. This additional question involves investigating:

- The relationship between the different measures, and interactions based on the different learning group.
- Learners' level of vocabulary size at the PY.

Similar to Chapter 4, the above items entail determining the initial levels of learners of the two learning groups at the beginning of their programmes (pre-test scores), their

knowledge at the end of the programme (pre-test scores), and the significance of change (i.e. improvement) in all measures, in both groups.

### 5.4.1. Effect of method of AWL learning on receptive and productive gains

*5.4.1.1. Effect of learning method on Receptive knowledge of the AWL.*

*Reliability and normality*

The items for the VLT (Receptive) test of receptive knowledge were checked for reliability and normality of distribution using *Cronbach's Alpha* test and *Shapiro-Wilk* test to confirm which statistical analysis would be performed with the data. In Table 5.4, below are the figures of reliability and normality distribution, and the histograms are illustrations:

**Table 5.4 Tests of normality and reliability of Receptive test of DG and NnDG groups:**

| Group | Rec test (VLT) | Cronbach's Alpha | Shapiro-Wilk | | | N |
|---|---|---|---|---|---|---|
| | | | Skewness | Kurtosis | sig | |
| DG | Pre | .770 | -.519 | -.735 | .007 | |
| | Post | .741 | -.956 | .961 | .016 | 44 |
| NnDG | Pre | .880 | -1.097 | .968 | .004 | |
| | Post | .915 | -1.005 | .351 | .001 | 43 |

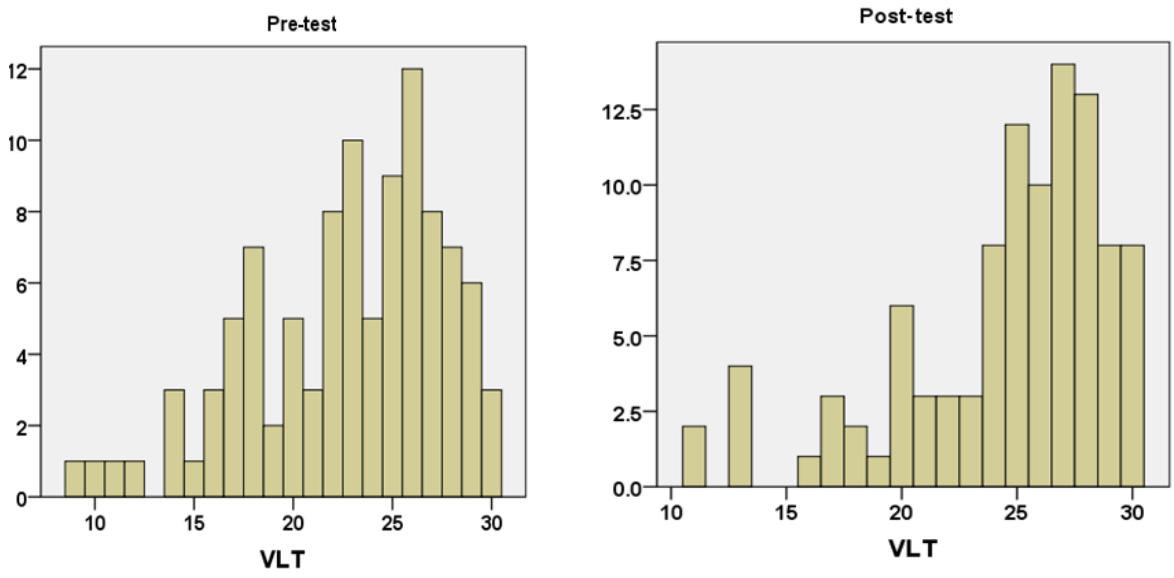



**Figure 5-1: Histograms of the AWL receptive test VLT**

The reliability of Cronbach's Alpha of the items of the VLT of the AWL was above 0.7. The normality figures of all scores in the pre and post-tests in both groups show that the data was not normally distributed, suggesting the use of the non-parametric analysis of the variances. The data demanded ranking the values of the test scores in order to perform a more robust analysis as will be discussed below. Similar to the *first study*, the reason for the abnormality of the distribution could be due to the ceiling effect toward the maximum score.

*The initial Receptive AWL knowledge of learners in both groups at Saudi PY:*

The scores of the receptive test show that both groups of advanced EFL Saudi adult learners have above moderate recognition levels of the AWL. They could recognize at least two thirds of the academic words. Learners from the treatment group or direct group DG show greater ability in this measure initially (=24/30) than the non-direct group NnDG (=21.2/30). The minimums and standard deviation proves this too. Based on the

formula set by Schmitt and AlHomoud (2007) discussed in section 4.5.1., the scores show that learners could recognize at least 403 academic head words.

**Table 5.5 Descriptive statistics of scores of VLT of DG and NnDG groups**

| Group | Pre Test | Mean /30 | ± SD | Minimum | Maximum | Range | *N* |
|---|---|---|---|---|---|---|---|
| DG | Rec or VLT | 24 | ±4.46 | 15 | 30 | 15 | 44 |
| NnDG | | 21.2 | ±6.06 | 4 | 30 | 26 | 43 |

*AWL receptive improvement over the PY programme:*

The general distributions of scores of both groups show a very weak improvement in receptive knowledge. The means of pre and post-test were almost the same as seen in Table 5.6 below. The results of the *first study* discussed the fact that learners were not expected to improve greatly since they showed higher marks initially on this measure. This claim became clearer since the lower achiever group of the *first study (SUni1)* showed higher improvement levels compared to *RUni2*. Concerning the data of this study (*the second study*); even the lower achiever group had a concrete sign of increase in the VLT score. The brief descriptive results are given in Table 5.6, below and illustrated in Figure 5.2.

**Table 5.6 Basic descriptive statistics of means and standard deviation of both tests of Receptive test (VLT)**

| Group | PreTest | ± SD | Post-test | ± SD | *N* |
|---|---|---|---|---|---|
| DG | 24 | ±4.46 | 24.3 | ±3.64 | 44 |
| NnDG | 21.2 | ±6.06 | 21.3 | ±6.81 | 43 |

Max = 30

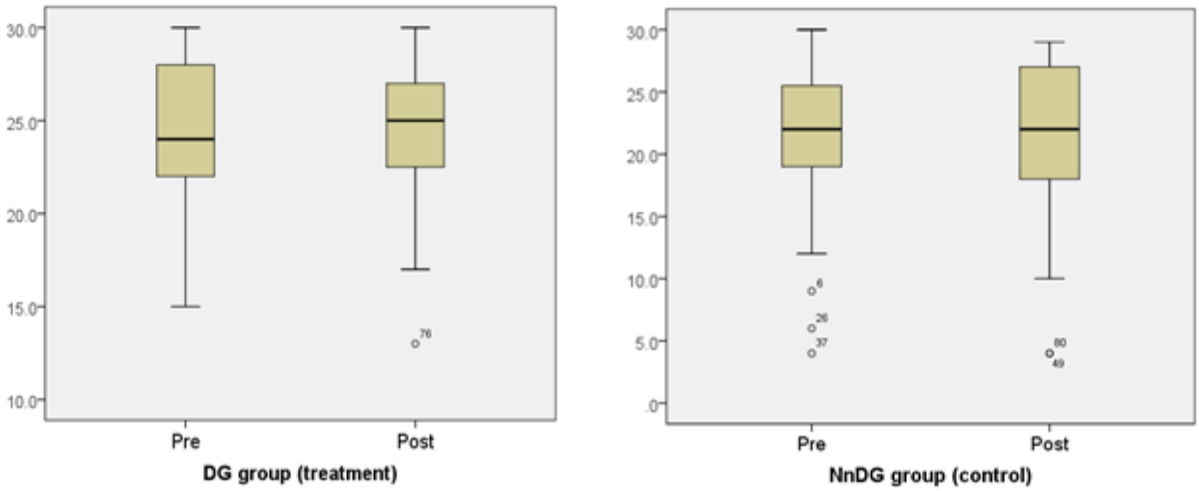



**Figure 5-2: improvement of receptive knowledge of the AWL (VLT test) according to groups**

The descriptive figures show no clear signs of improvement in either group. As argued, based on the previous study concerning learners with low pre scores in this measure, the learners of NnDG were expected to show some level of improvement. Furthermore, this study hypothesized that learners from the treatment group would be expected to gain some academic words due to the extra activities they were exposed to. No clear difference between pre and post scores was found in this measure.

Looking at the table, a slight pre-post difference can be noticed; however, the initial analysis could not capture the difference. A further different analysis was performed and confirmed the significant difference. This analysis was performed by running the *Wilcoxon test* due to the abnormality of the distribution of the data. The comparisons of the DG (or treatment group) and NnDG (or control group) show no significant pre-test/post-test differences at ($p>.05$). This analysis was performed by running the non-parametric paired test on the original data. A possible justification for the insignificance of the improvement is that learners showed higher levels of receptive knowledge, and therefore were not likely to have a notable increase. Also, based on the results of the *first*

*study*, the increase in learners' productive knowledge, (scores of CATSS), was higher and more obvious than their performance in receptive knowledge as reported in Chapter 4. This was also true for participants of this study as seen below in the following section (i.e. learners showed significant improvement in CATSS in this study, see below).

Nonetheless, as discussed, this study hypothesizes that a variation must have occurred between the two groups due to the different AWL input they were exposed to, and that at least the treatment group DG would have improved. Therefore, a further more robust analysis was performed to confirm any significant difference in terms of improvement from pre-test to post-test in this measure. The test is the nonparametric *Levene's tests* that is based on ranked data (Nordstokke & Zumbo, 2007). It is a traditional *Levene's test*, but, the original scores are given ranked values, and then the mean is calculated then subtracted from the ranked scores. This test showed significant difference in terms of improvement from pre-test to post-test for the DG (or treatment group), and showed no significance improvement in the NnDG (or control group). Table 5.7 bellow shows the figures regarding significance of improvement, medians, minima and maxima.

**Table 5.7 Descriptive statistics of REC (VLT) scores of Direct Group DG and Non Direct Group NnDG:**

| School | Test | Mean +-SD 30/30 | Diff | P Value | *Min* | *Max* | *Median 30/* | *N* |
|---|---|---|---|---|---|---|---|---|
| DG | RecPre | 24.0± 4.4 | 0.3 | .039 [a] | 15 | 30 | 24 | |
| | RecPost | 24.3± 3.6 | | .041 [b] | 13 | 30 | 25 | 44 |
| NnDG | RecPre | 21.2 ± 6.0 | 0.1 | .231 [a] | 4.0 | 30 | 22 | |
| | RecPost | 21.3 ± 6.8 | | .257 [b] | 4.0 | 29 | 22 | 43 |

a based on the non-parametric *Levene's test* on ranked data describe above.

b based on the *Wilcoxon test* of the ranked data.

The above analysis of difference shows a significant improvement in the learners of the treatment group. Learners from the control group who received normal academic English teaching did not show any evidence of improvement. This clearly suggests that the direct enhancement of academic vocabulary for advanced learners impacted more positively on learners than learning the AWL within the contexts of 'normal' teaching of academic English only. In other words, the advanced learners in the passive group (NnDG), who did not participate in any direct extra activities which enhanced the AWL, failed to show any progress in their receptive knowledge. The academic English they were exposed to was not sufficient to result in an increase in receptive knowledge of the AWL. The capability of those advanced learners might not be sufficient to allow them to acquire more academic words, depending only on academic reading and listening in their advanced English programme.

Nonetheless, the increase of the AWL observed in the DG group may be considered unsatisfactory. The rate of improvement depending on the median scores was (1) word, which is equal to 19 words based on the formula used in Section 4.5.1.

There are further discussions about the implications of these results in Chapter 7.

#### *5.4.1.2. Effect of learning method on Productive knowledge of the AWL. Reliability and normality*

The items of the productive measure of academic words CATSS for both groups were checked for reliability and normality of distribution in order to confirm which statistical analysis should be performed with the data. The figures are shown in Table 5.8. below:

**Table 5.8 Tests of normality and reliability of productive tests of DG and NnDG groups:**

| Group | CATSS | Cronbach's Alpha | Shapiro-Wilk | | | N |
|---|---|---|---|---|---|---|
| | | | Skewness | Kurtosis | Sig | |
| DG | Pre | .790 | .361 | -1.011 | .021 | 44 |
| | Post | .841 | .065 | -1.168 | .025 | |
| NnDG | Pre | .821 | .297 | -.977 | .063 | 43 |
| | Post | .903 | .623 | -.857 | .002 | |

The reliability of Cronbach's Alpha of all items of CATSS, of both tests of each group, showed a strong reliability figure at more than (.790). Regarding the normality of distribution of responses, all test items apart from the pre-test items of NnDG were abnormal, suggesting the use of non-parametric analysis. The reason for the non-normal of distribution could be due to the clear skewness towards the minimums. The graphical illustration is displayed below in Figure 5.3:

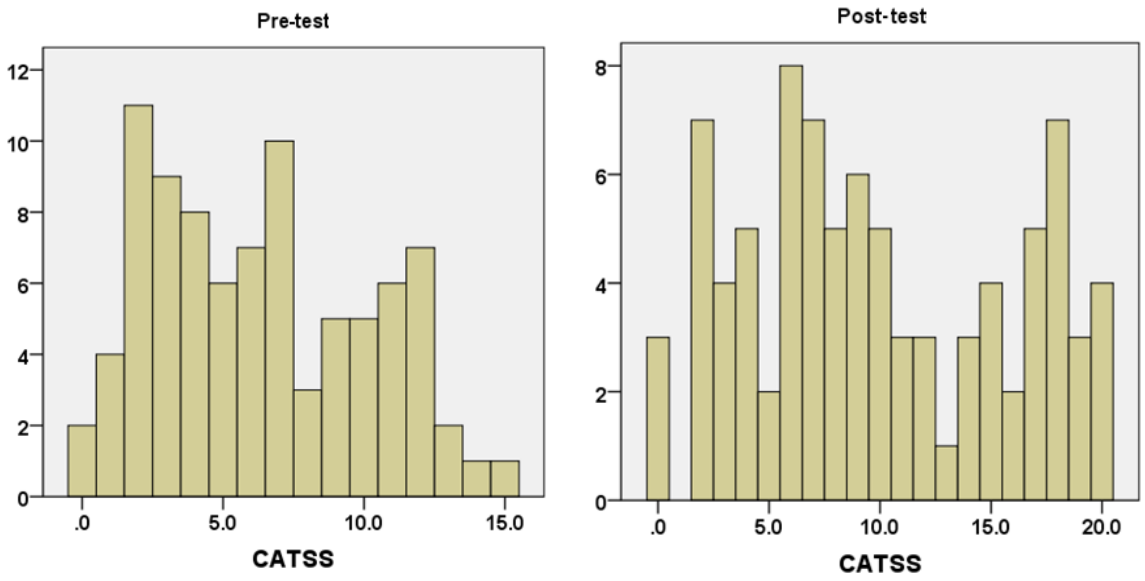



**Figure 5-3: Histograms of the AWL receptive test CATSS**

*The initial productive AWL knowledge of learners in both groups at Saudi PY:*

The scores of the productive knowledge test *CATSS* show that both groups of Saudi advanced PY learners have a poor ability in producing the AWL. Learners from both groups seem to have had the same level of productive knowledge of academic words at the beginning of the term. They shared almost the same productive knowledge figures initially, as seen in Table 5.9. below. This level of productive knowledge of the AWL is equal to 114 academic words, based on the same formula of estimations discussed in the previous sections. The maximum scores show that even the high achievers among advanced Saudi EFL learners at the PY could hardly produce 50% of the academic list. Table 5.9 below shows the pre-test mean, the standard deviation, the minimum, maximum and range scores of the productive test CATSS.

**Table 5.9 Initial descriptive statistics of scores of CATSS of DG and NnDG groups:**

| Group | Pre Test | Mean /30 | ± SD | Minimum | Maximum | Range | *N* |
|---|---|---|---|---|---|---|---|
| DG | CATSS | 6.66 | ±3.80 | 1 | 15 | 14 | 44 |
| NnDG | | 6.00 | ±3.83 | 0 | 14 | 14 | 43 |

*The AWL productive improvement over the PY programme:*

The general descriptive figures of scores of both groups of learners show a clear improvement in productive knowledge. This was predictable in this type of measure, since learners showed poorer gains in the pre-tests. However, what was interesting was that the treatment group DG clearly improved more than the passive group NnDG, despite showing similar levels of knowledge initially. The descriptive figures of the improvement can be seen in Table 5.10, and this is illustrated in Figure 5.4.

**Table 5.10 Basic descriptive statistics of means and standard deviation of both tests of CATSS**

| Group | PreTest | ± SD | Post-test | ± SD | *N* |
|---|---|---|---|---|---|
| DG | 6.659 | ± 3.80 | 11.727 | ± 5.16 | 44 |
| NnDG | 6.000 | ± 3.83 | 8.023 | ± 6.02 | 43 |

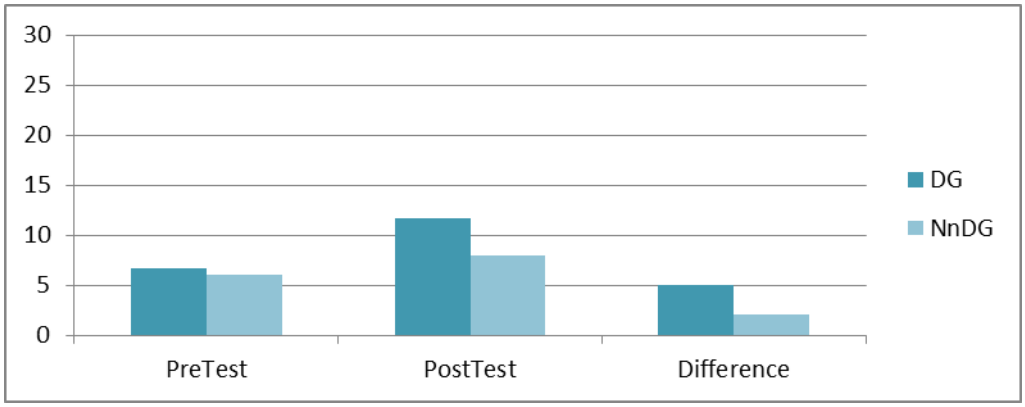



**Figure 5-4: Histograms of improvement of in the productive knowledge CATSS of Saudi PY**

Analysis of comparisons confirmed the significance of difference (i.e. pre-test/post-test improvement), between learners of both groups. The analysis was performed by running the *Wilcoxon test* on both groups due to the abnormality of the distribution of the data. The analysis showed a significant difference in terms of improvement from pre-test to post-test for both groups; DG (or treatment group) at ($p<.000$) and a significance figure with NnDG (or control group) at ($p<.004$). Table 5.11 below shows the figures regarding significant of improvement, medians, minimums and maximums.

**Table 5.11 Descriptive statistics of *Productive* scores of Direct Group DG and Non Direct Group NnDG:**

| School | Test CATSS | Mean +-SD 30/30 | Diff | P Value | *Min* | *Max* | *Median 30/* | *N* |
|---|---|---|---|---|---|---|---|---|
| DG | Pre | 6.659 ± 3.80 | 5.068 | .000 | 1.0 | 15.0 | 6.00 | |
| | Post | 11.727± 5.16 | | | 2.0 | 20.0 | 10.15 | 44 |
| NnDG | Pre | 6.000 ± 3.83 | 2.023 | .004 | 0 | 14.0 | 6.00 | |
| | Post | 8.023 ± 6.02 | | | 0 | 20.0 | 6.00 | 43 |

Based on the non-parametric *Wilcoxon test*.

The detailed statistics confirm the improvement by all learners regarding productive gains of the academic words during the term (12 weeks). There is a significant mean difference between pre-test and post-test for CATSS. The degree of improvement in the passive group NnDG is lower than that of the treatment group. The pre/post mean difference was lower, and the median remained the same as in the pre-test. Statistically, in terms of the difference regarding the degree of improvement between the two groups, the data was analysed using the two-way *ANOVA* and showed that the improvement in the DG (active group) was significantly higher than that of the NnDG (passive) group: Wilks' $\lambda = .880$ $F(1, 85) = 11.629$, $P < .001$. The line chart in Figure 5.5 below graphically illustrates the variation in terms of level of improvement.

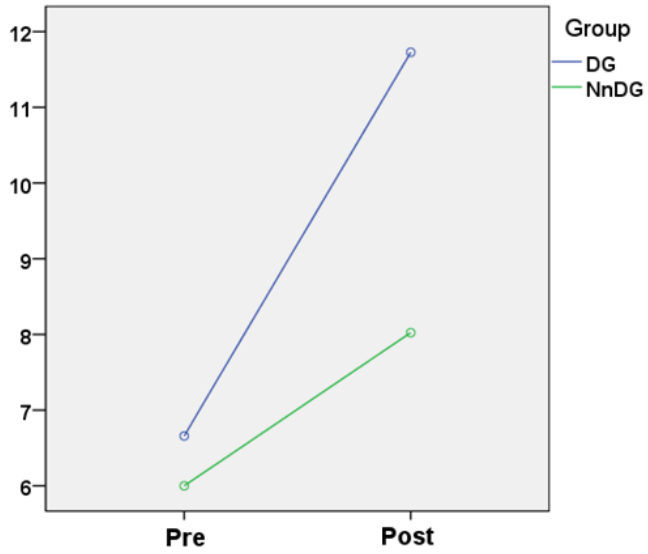



**Figure 5-5: Line Chart: Pre/Post-tests according to groups**

The extra direct enhancement of activities that targeted the AWL impacted more positively on learners than learning the AWL incidentally within academic English teaching. Although learners had an advanced level of English proficiency, and were perhaps considered capable of learning language input incidentally, learners of the

passive group showed a weaker increase with regard to the AWL in an academic English programme. In fact, though their improvement was statistically significant, it was not particularly strong as their counterparts of the treatment group, based on the *ANOVA* reported above. This indicates the value of focusing on lexical teaching of academic words to advanced learners, even if they show potential ability to learn language input incidentally. The common conception discussed earlier in the thesis that favours the incidental learning of language input with advanced students due to their ability, might not always be reliable. Although the learners in this study showed almost the same levels of AWL productive knowledge in the pre-test, and received normal academic English teaching, only the learners from the DG (treatment group) increased their range of vocabulary significantly higher.

In real terms, increased knowledge of academic words by learners of the DG group was equal to 96.4 words, based on the formula of estimation mentioned earlier. The NnDG gained about 38 words.

#### *5.4.1.3. Advanced learners overall vocabulary size in the Saudi PY.*

As discussed earlier in this section, the general vocabulary size is explored to respond to the sub-research question and the research hypotheses of this study. For the sake of organization, the relationship between vocabulary size and vocabulary knowledge based on the groups is discussed later in 5.4.1.5. The results regarding this question are briefly reported below in Table 5.12 and 5.13.

*Reliability and normality*

The same procedure concerning reliability and normality check was performed on the data of the XKlex test of vocabulary size. The test versions were shown as reliable, with alpha estimates of .882 in the pre-test and post-test of the DG, and .918 in the pre-test and post-test of the NnDG. Regarding the normality of distribution, the data of all items of both groups was normal at (sig >.05).

*Vocabulary size of PY learners, and their improvement*

The two groups of PY learners gave two different vocabulary size score estimates based on version (A) of XKlex. The learners of the treatment group DG had a larger vocabulary size initially in the term of the PY at 5965 words. Learners from the NnDG group had an estimate of 4809 words.

This is illustrated in Table 5.12, which also indicates the SD among learners:

**Table 5.12 Descriptive statistics of scores of DG and NnDG of the PY**

| | School | Mean /10K words | ± SD | Minimum | Maximum | *N* |
|---|---|---|---|---|---|---|
| XKlex pre scores | DG | 5965 | ±1347 | 200 | 8300 | 44 |
| | NnDG | 4809 | ±1440 | 60 | 7700 | 41 |

It was argued above, and in the *first study* that learners are not expected to expand their vocabulary size greatly in a limited time, because the development of general vocabulary is a slow process. As stated in the *first study*, the post-test could be used at least as an estimate of learners' vocabulary size, since XKlex has been criticized for giving on overestimation of vocabulary size due to the way it works (see the previous discussion in section 3.4.1.3.). However, as reported in the *first study*, participants (n= 83) increased their range of vocabulary by about 390 words during the presessional courses.

Surprisingly, the current study revealed unexpected results regarding the estimates of the post-tests. The means of the post-scores of learners of both groups had in fact decreased since the pre-tests. The estimates of scores of version (B) of XKlex decreased by 534 words in DG, and 1029 words in NnDG. The comparative analysis was performed using the *paired sample t-Test.* The difference was significant at (<.001). The figures in Table 5.13 below show a detailed analysis of the results of the comparison:

**Table 5.13 Descriptive statistics of score comparisons of DG and NnDG**

| School | XKlex | Mean +-SD 10k | Diff | P Value | *Min* | *Max* | *Median 30/* | *N* |
|---|---|---|---|---|---|---|---|---|
| DG | Pre | 5965±1347 | -534 | .001 | 2000 | 8300 | 5950 | |
| | Post | 5431±1297 | | | 1200 | 7700 | 5600 | 44 |
| NnDG | Pre | 4809±1440 | -1029 | .000 | 600 | 7700 | 5100 | |
| | Post | 3780±1023 | | | 1700 | 6700 | 3600 | 41 |

Based on *paired sample t-Test* analysis.

The results in Table 5.13 above show a significant decrease in the post scores of XKlex in both groups. The NnDG decrease is nearly double that of the DG. The decrease was also clear in medians, maxima, and was available in minima scores.

It was initially suspected that the reason for the unexpected decrease might be due to the modification of version (B) of Xklex discussed earlier in procedures (section 5.3.), and at the end of the *first study* (section 4.6.). i.e. the construction of version (B) of the XKlex test was modified to lose the pattern of *non-words* that the authors of the test had created in their test. Although XKlex test was extensively validated by Masrai (2009), it was decided in this study to change the pattern of the *non-words,* since Saudi learners tend use a lot of guess work in answering unknown words (AlHazemi's, 1993 *cf* Masrai, 2009). It was decided at the time of the study to have no pattern in the items on one test

not on both of them (i.e. either pre or post). The reason was that the study follows the same procedure of (Masrai, 2009) on one occasion (i.e. either pre- or post-test), and the pattern is then removed due to its potential risks on the other occasion. Losing the pattern on both test occasions might result in discomfort in claiming the estimates of learners' lexical size. At any rate, learners' vocabulary size was not expected to expand greatly during a term, as argued previously.

Nonetheless, there was no certainty that learners had performed better in the pre-test than in the post-test because of the pattern of the items. Furthermore, there was no evidence in participants' test sheets to show that the pattern of version (A) was figured out by learners. Added to this, when looking at the highest values in learners' answers in both groups in the pre-test, there was no score that would be considered abnormal. i.e. if learners have figured a pattern, they are more likely to reach the maximum test score. There was no skewness as reported earlier.

However, as argued above, the unexpected decrease caused suspicion. Thus, a limited follow-up study was conducted in order to have more stable grounds concerning the generalization of the post scores. It was intended to explore whether Saudi participants used guesswork in answering the version of XKlex which contained the pattern.

#### *5.4.1.4. A Follow up study: Learners' feedback on the construction of XKlex test.*

*Aim of the study:*

The aim of this limited follow-up study was to explore Saudi test takers' behaviour regarding the XKlex test concerning the pattern in its construction. The study intended to investigate learners' reaction towards the test and whether they used a lot of guesswork, which would thus influence their responses. The main focus of the study was whether the

pattern in the construction of the XKlex was easily detectable or not. It was hoped that the results of the study would confirm whether learners' high level performance in the pre-test represents an accurate estimate of their vocabulary size, and that there was no overestimation influenced by the pattern.

*Procedure*:

10 Saudi postgraduate learners participated in this study. They were administered version (A) and version (B) of XKlex (Masrai, 2009) one immediatedly after the other. There was no modification to the test, and both versions had the original pattern created by the author (see section 3.4.1.). Learners were then asked to answer four short questions to express their thoughts about the test. The key question in this short questionnaire was whether or not they had noticed a pattern in the construction of the test. However, to guarantee more reliability concerning learners' reaction to this question, the questionnaire was designed in a particular way to help reach conclusions about their feedback indirectly. It was thought at the time of the study that direct questions might not lead to expressions of their beliefs accurately, leading to doubts about participants' actual responses; i.e. some learners might deny knowing the pattern in order to please the researcher, or others might claim knowing a pattern for competitive reasons with their classmates.

Thus, the questions were written individually on small cards and were given one after the other to students after they had finished the XKlex tests. The wording of the questions started vaguely and moved into being more precise. The Arabic translation was provided beside each question. Table 5.14 below lists the questions:

**Table 5.14 Questions of the short questionnaire**

| | |
|---|---|
| Briefly, how do you describe this test? باختصار, كيف تصف هذا الاختبار<br>……. | 1 |
| Do think the test is easy? هل باعتقادك ان الاختبار كان سهلا؟<br>……. | 2 |
| Do you think that the construction of the test has something that made it easier?<br>هل تعتقد ان هناك شيئا ما بتصميم الاختبار جعله سهلا؟<br>……. | 3 |
| Did you notice a pattern in the test that helped you with the answers, and if so, what was it?<br>هل لاحظت ان عناصر الاختبار منسقة بشكل معين ساعدك على الحل, وما هو؟.<br>……. | 4 |

*The results of this limited study*:

It could be said confidently from 10 Saudi participants that none of them had worked out a pattern in either version of Xklex. In fact, their comments in question 4 which was meant to come last in the questionnaire and asked directly about the pattern, indicated that they had no clue about any pattern in the construction. Some learners claimed to know a pattern, but it was not the pattern intended in the test, and they were referring to the non-words, as seen below. One of participants' comments (ID 10) was not clear regarding confirming awareness of the pattern, as inferred from his comment mentioned below. However, the real reason for the test was always explained to all of the testees at the end of the questionnaire, and none of them declared knowing the pattern. Besides, if student (10) had noticed, we would have expected that he would have scored higher, which was not the case. The table below provides their answers to the short questionnaire. Some of participants' actual words in the comments were edited, especially in the warm-up question.

**Table 5.15 Answer to the short questionnaire**

| Parts' ID | Describe this test. | Was test easy? | Construction made easy? | Did you notice a pattern? |
|---|---|---|---|---|
| IZ | Interesting | Yes but not always | Yes | 'Yes, some words were not English but they seemed so. |
| 2 | described the test | No, not sure about spelling | Word order not clear for me | No |
| WR | Good to test my vocabulary | Yes for the words I know | No | 'group easy to difficult' 'some words were not right' |
| BS | Strange test, checking common words | Yes | Yes, clear tables | 'No' |
| WS | Good in some words | Easy in some words | I don't think that | 'No' |
| IM | It's spelling vocabulary test | Yes | Yes I do | 'No I don't' |
| FM | Checking spelling | Yes | Maybe | 'I didn't notice any' |
| SD | Vocabulary knowledge test | kind of | Yes | 'No' |
| MD | Easy but not sure of some words | Not sure but looked easy | Yes, ticking made it easy | 'Nothing,' 'some words weren't real' |
| 10 | Clear | Easy for the words I know | Dividing the words … | 'Yes, it contained dividing words, I knew immediately' |

### *5.4.1.5. The relations between vocabulary size XKlex and the knowledge of the academic words in CATSS and VLT.*

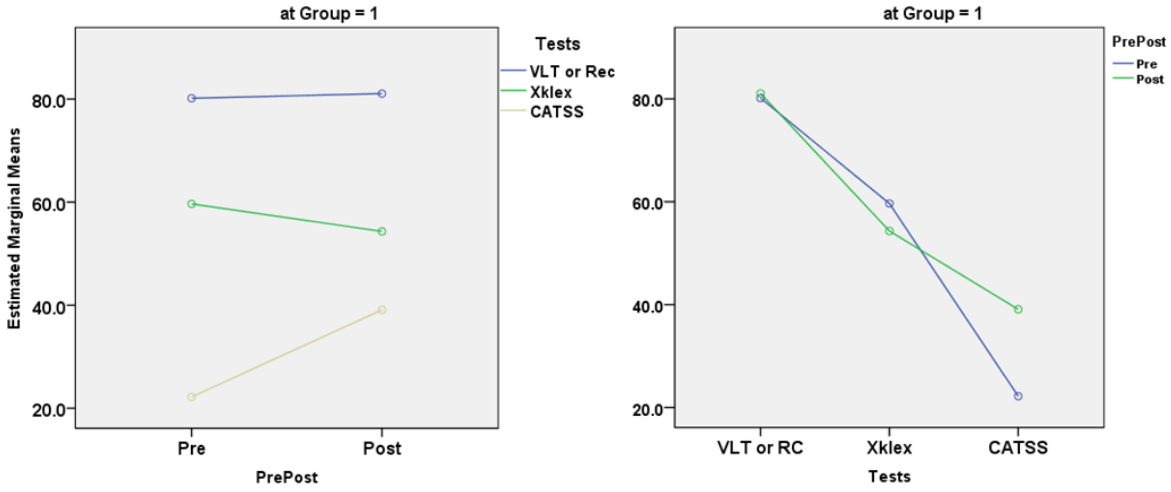
To answer these two questions, the scores from the three different measures (CATSS, VLT and XKlex) in both learning groups were used. Similarly to the *first study*, the original scores of all tests were transformed into percentages since each test had a different maximum as explained in procedures. The descriptive statistics in Table 5.16, and 5.17. below explain the variations between the tests in each group with scores transformed into percentages, and these are also graphically illustrated in Figures 5.3. and 5.4. below.

**Table 5.16 Descriptive statistics of all scores of participants of DG transformed to 100%**

| | Test | original score | Mean | SD | *Min* | *Max* | *N* |
|---|---|---|---|---|---|---|---|
| Pre | CATSS | /30 | 22.19 | 12.67 | 3.00 | 50.00 | 44 |
| | VLT or REC | /30 | 88.15 | 14.88 | 50.00 | 100.00 | |
| | XKles | /10k | 59.65 | 13.47 | 20.00 | 83.00 | |
| Post | CATSS | /30 | 39.09 | 17.22 | 6.00 | 66.66 | 44 |
| | VLT or REC | /30 | 81.06 | 12.15 | 43.33 | 100.00 | |
| | XKles | /10k | 54.31 | 12.97 | 12.00 | 77.00 | |

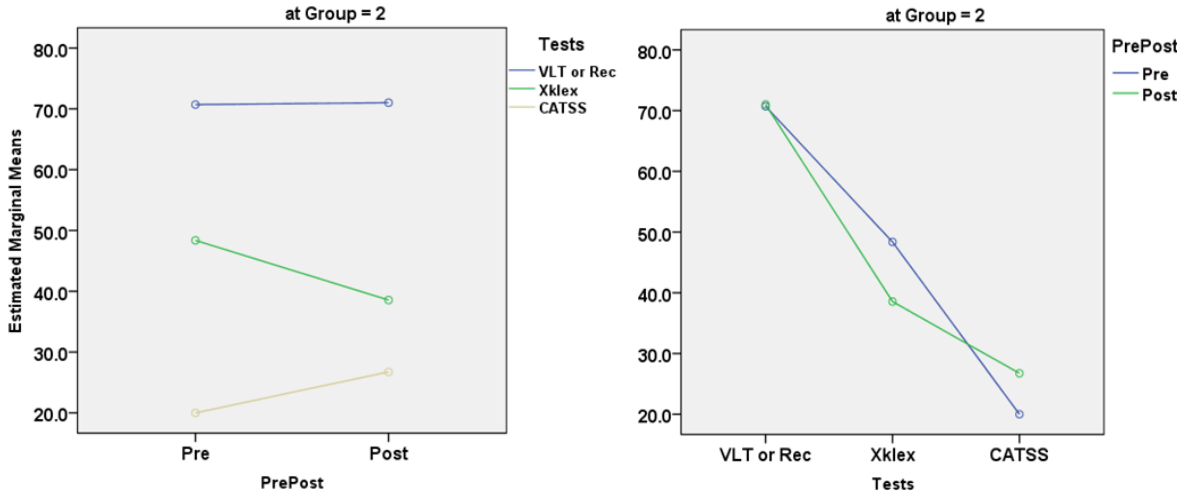
**Table 5.17 Descriptive statistics of all scores of participants of NnDG transformed to 100%**

| | Test | original score | Mean | SD | *Min* | *Max* | *N* |
|---|---|---|---|---|---|---|---|
| Pre | CATSS | /30 | 20.24 | 12.91 | 0.00 | 46.66 | 41 |
| | VLT | /30 | 71.86 | 18.54 | 20.00 | 100.00 | |
| | XKles | /10k | 48.09 | 14.40 | 6.00 | 77.00 | |
| Post | CATSS | /30 | 27.23 | 20.09 | 0.00 | 66.66 | 41 |
| | VLT | /30 | 72.19 | 21.32 | 13.33 | 96.66 | |
| | XKles | /10k | 37.80 | 10.23 | 17.00 | 67.00 | |




***development and variations*** ***interaction***

**Figure 5-6: Pre-test Post-test variations and interactions of DG tests**




***development and variations*** ***interaction***

**Figure 5-7: Pre-test Post-test variations and interactions of NnDG tests**

The data in the tables and figures of both groups show significant variations between the means. As discussed previously in the *first study*, the variation between learners' performance in CATSS and VLT was obviously expected since they tend to assess two different AWL mastery abilities, and both of them vary from XKlex as it measures overall vocabulary size (including the AWL).

The figures show clearly some variations in improvement between learners in the treatment group DG, and learners in the passive group NnDG. The increase in VLT (receptive knowledge) of the DG was slightly more obvious than in the NnDG. The increase in CATSS (productive knowledge) was clearer in both groups, but was obviously greater in the DG, as discussed also previously in sections 5.4.1.1 and 5.4.1.2. The variation between learners' performance in both groups is discussed further in Chapter 7.

The data was analysed by the two-way mixed design *MANOVA* based on the ranked data to avoid violations regarding normality (Beasley & Zumbo, 2009; Field, 2013) and

showed significant difference between learners' performances in the different tests with respect to their two group of learning F (2, 79) = 4.499, *P<0.05*.

Multivariate tests indicated that there was an effect based on group (i.e. AWL enhancement or treatment effect) on the variation in the pre-tests/post-tests improvements of the different measures;

*All Tests Pre & Post * Group: F (6, 338) = 14.339, P<0.00.* The test between subject effects is illustrated in the Table 5.18 below.

**Table 5.18 Interactions between subject effects**

| | | | *F* | Sig |
|---|---|---|---|---|
| PrePost * Group | ↔ | VLT | 6.479 | .002 |
| | ↔ | Xklex | 31.511 | .000 |
| | ↔ | CATSS | 11.641 | .000 |

In terms of improvement, the data showed significant difference between the pre-tests/post-tests in the DG group F (2, 42) = 29.419, P>.000, and in NnDG group F (2, 36) = 6.835, P>.003.

Note that among the tests, scores of XKlex decreased in the post-test as discussed in the previous section, which therefore had an effect on the figures. All the data transformed into percentages is analysed by *Wilcoxon Signed Ranks Test* to indicate the pairwise differences (separate Ttest analyses or their nonparametric counterparts were discussed above in the results within the different sections above). The pairwise tables are shown below, plus the significance of interaction for both groups.

***Direct Group DG:***

**Table 5.19 Pairwise statistics of all measures**

| | Pre | Post | Mean Difference | *T* | Sig. (2-tailed) |
|---|---|---|---|---|---|
| CATSS | 22.19 | 39.09 | 16.09 | -5.177 | .000 |
| VLT | 80.15 | 81.06 | .91 | .264 | .621[a] |
| XKles | 59.31 | 54.31 | -5.34 | -2.702 | .007 |

[a] *An alternative test shown significance as reported earlier in section 5.4.1.1*

**Table 5.20 Interactions between all measures**

| | Sig |
|---|---|
| CATSS ↔ VLT: Pre*Post | .000 |
| CATSS ↔ XKlex: Pre*Post | .000 |
| VLT ↔ XKlex: Pre*Post | .024 |

***Non- Direct Group NnDG:***

**Table 5.21 Pairwise statistics of scores of all measures**

| | Pre | Post | Mean Difference | *T* | Sig. (2-tailed) |
|---|---|---|---|---|---|
| CATSS | 20.24 | 27.23 | 15.87 | -2.905 | .004 |
| VLT | 71.86 | 72.19 | .33 | -.330 | .741[a] |
| XKles | 48.09 | 37.80 | -10.29 | 3.916 | .000 |

[a] *An alternative test shown also no significance as reported earlier in section 5.4.1.1*

**Table 5.22 Interactions between all measures**

| | Sig |
|---|---|
| CATSS ↔ VLT: Pre*Post | .000 |
| CATSS ↔ XKlex: Pre*Post | .000 |
| VLT ↔ XKlex: Pre*Post | .005 |

## *Correlation between different abilities:*

The two groups had different correlation figures in the three tests in both pre and post-tests. In general, the different measures correlate significantly with each other in pre-test and post-test in the DG group. Concerning the NnDG group, the correlation was only clearer between the two measures of the AWL knowledge, than with XK-lex. The Tables below summarize the correlations between all measures in the pre and post-tests of both groups:

***Direct group DG:***

**Table 5.23 correlation figures of DG *pre test***

| | | N | CATSS | VLT | XKlex |
|---|---|---|---|---|---|
| CATSS | Correlation Coefficient | 44 | 1.000 | .650** | .466** |
| | Sig. (2-tailed) | | | .000 | .001 |
| VLT | Correlation Coefficient | | . 650** | 1.000 | .378* |
| | Sig. (2-tailed) | | .000 | | .011 |
| XKlex | Correlation Coefficient | | .466** | . .378* | 1.000 |
| | Sig. (2-tailed) | | .001 | .011 | |

**Table 5.24 correlation figures of DG *post-test***

| | | N | CATSS | VLT | XKlex |
|---|---|---|---|---|---|
| CATSS | Correlation Coefficient | 44 | 1.000 | .542** | .503** |
| | Sig. (2-tailed) | | | .000 | .001 |
| VLT | Correlation Coefficient | | .542** | 1.000 | .537** |
| | Sig. (2-tailed) | | .000 | | .000 |
| XKlex | Correlation Coefficient | | .503** | .537** | 1.000 |
| | Sig. (2-tailed) | | .001 | .000 | |

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

***Non- Direct group NnDG:***

**Table 5.25 correlation figures of *NnDG pre test***

| | | N | CATSS | VLT | XKlex |
|---|---|---|---|---|---|
| CATSS | Correlation Coefficient | 43 | 1.000 | .459** | .127 |
| | Sig. (2-tailed) | | | .002 | .422 |
| VLT | Correlation Coefficient | | .459** | 1.000 | .380* |
| | Sig. (2-tailed) | | .002 | | .013 |
| XKlex | Correlation Coefficient | | .127 | .380* | 1.000 |
| | Sig. (2-tailed) | | .422 | .013 | |

**Table 5.26 correlation figures of *NnDG post test***

| | | N | CATSS | VLT | XKlex |
|---|---|---|---|---|---|
| CATSS | Correlation Coefficient | 43 | 1.000 | .500** | .422** |
| | Sig. (2-tailed) | | | .001 | .000 |
| VLT | Correlation Coefficient | | .500** | 1.000 | .162 |
| | Sig. (2-tailed) | | .001 | | .304 |
| XKlex | Correlation Coefficient | | .422** | .162 | 1.000 |
| | Sig. (2-tailed) | | .000 | .304 | |

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed)

Concerning the treatment group DG, learners' CATSS and VLT (receptive and productive) scores were found to correlate positively, but more highly in the pre-test than the post-test. Some correlation was found between the knowledge of the academic words (productively and receptively) and overall vocabulary size. The scores of CATSS was found to correlate moderately with XKlex in the pre and post (.466** and .503** respectively). The correlation between VLT and XKlex was weak in the pre-test and moderate in the post-test.

Learners of the non-direct group NnDG showed moderate correlation between productive AWL knowledge and lexical size. The correlation in the pre-test was slightly lower than in the post-test. The correlations between the two AWL knowledge abilities (productively and receptively), and the lexical knowledge were weaker and less significant. CATSS and XKlex showed weak and insignificant correlation in the pre-test and moderate in the post-test, and the VLT had modest correlation with XKlex in the pre-test and weaker and insignificant correlation in the post-test.

As expected, there was a higher positive correlation between CATSS and VLT since they are two measures assessing the same area, i.e. academic words. The XKlex correlation with the two AWL measures was also expected since the test includes the AWL within its frequency bands. However, the correlation was clearer and higher in the DG group than in the NnDG. Most probably the reason was the unexpected decrease in XKlex that was discussed in section5.4.1.3.

### 5.4.2. Effect of method to AWL learning on lexical richness of the AWL in writing.

To address this hypothesis, the following were identified: the initial AWL lexical frequency in learners' writing at the beginning of the term for both learning groups (DG and NnDG), and the level of the AWL lexical frequency profile in their writing at the end of the term. The relationship and correlation between the improvement of the AWL (receptively and productively) by writers from the two groups and their lexical academic profile in writing is also addressed in section 5.4.3.

Unlike the *first study* where the written data were collected in class on two occasions, this study used written data from two writing tests during the term. As mentioned in the procedures, the data of only 6 students (4 from DG and 2 from NnDG group) qualified to be used for the analysis (further data from NnDG was used in section 5.4.2.3.). The data was typed into *MS Word processer* and then run through the *vocabprofiler*. It was stated in the procedure that the misspelled or unclear AWL words from learners' writing were shown to a native speaker validator to be corrected into proper words before processed into *vocabprofiler.* Only 2 academic words needed correction. The original words were copied into two documents; one was corrected provisionally, and the other was submitted to the validator for correction. The agreement with validator was 100% in this study.

#### *5.4.2.1. Descriptive reports from the vocabprofiler on learner' writing from both groups at the beginning of the term:*

Similarly to the *first study,* due to the limited data and fewer participants, it was thought that presenting the entire data of each participant in more details would be more useful.

Qualitative reports from the *vocabprofiler* of all participants at the beginning of the term in both groups are shown in Tables 5.27 and 5.28 below:

**Table 5.27 Descriptive reports from the vocabprofiler of learners from treatment group DG**

| Participants | Word Count | 1k+2k % in Text | AWL | | | | |
|---|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percent/AllWards | Percent/150Wds |
| 210 | 333 | 89.49% | 10 | 11 | 13 | 3.90% | 3.33% |
| 222 | 347 | 90.49% | 10 | 11 | 16 | 4.61% | 2.65% |
| 232 | 209 | 93.30% | 1 | 1 | 2 | 0.96% | 0.67% |
| 234 | 166 | 91.57% | 5 | 5 | 11 | 6.63% | 6.67% |

**Table 5.28 Descriptive reports from the vocabprofiler of learners from the passive group NnDG**
Definitions of families, tokens, and types are found in 2.4.4.

| Participants | Word Count | 1k+2k % in Text | AWL | | | | |
|---|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percent/AllWards | Percent/150Wds |
| 315 | 219 | 92.70% | 6 | 6 | 8 | 3.65% | 5.23% |
| 320 | 132 | 93.18% | 4 | 5 | 5 | 3.79% | 3.79% |

The two reports give information about learners' lexical frequencies in writing. The tables report how many words writers produced, then give figures about percentages of the general words (GSL 2000), and the academic words (AWL) based on both entire word count and on *150 words* only. The count of the AWL word families, types and token is also given. In the *first study*, the discussion of results included an analysis of writers based on the whole length of compositions, and on 200 words of all learners. In this study, the emphasis of the discussion of the results is mainly on the analysis based on the entire word length of compositions (i.e. all words written by learners in the compositions).

Based on the academic task, the initial AWL frequencies in the writing of 5 out of 6 learners at the beginning of the term was above 3.60%. According to the 5% figure claimed by Morris and Cobb (2004) discussed earlier, this is almost an acceptable figure

in academic writing. There was one lower achiever in the DG group who initially used 1 academic word in a 209 word composition. Apart from this participant, learners produced usually more than 5 AWL word types (create and creative are two *types* of one *family* word). The entire AWL *families* that each writer mentioned in their essays are available in Appendix H.

### *5.4.2.2. Descriptive reports from the vocabprofiler on learner' writing of both groups and their AWL frequencies at the end of the term:*

It was mentioned in the procedures that the only data that was found to be suitable for comparisons involved, unfortunately, an academic writing task initially and a less advanced (i.e. general English level, non-academic) writing task in the post-writing. Therefore, no reliable or meaningful comparison was possible. Nonetheless, if the difference in the genre of the writing task was disregarded, as well as the effect of small size population complications, the obvious result regarding comparisons in improvement between the two learning groups is that learners of the treatment group DG have produced some academic words in the second task, whereas the NnDG learners produced none. Unfortunately, this assumption is uncertain due to the limited data. Further reports and interaction graphs between measures of those individual learners may give some indications to support this claim to a greater extent as seen below in the following section. Tables 5.29 and 5.30 below summarize the AWL frequencies of the learners in both groups.

**Table 5.29 Descriptive reports from the vocabprofiler of the post-writing of learners from DG**

| Participant ID | Word Count | 1k+2k % in Text | AWL | | | | |
|---|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percent/AllWards | Percent/150Wds |
| 210 | 298 | 91.27% | 7 | 7 | 7 | 2.35% | 2.63% |
| 222 | 262 | 93.13% | 4 | 5 | 6 | 2.29% | 2.67% |
| 232 | 125 | 92.80% | 1 | 1 | 3 | 2.40% | 2.40% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 234 | 152 | 88.81% | 3 | 3 | 3 | 1.97% | 1.97% |

**Table 5.30 Descriptive reports from the vocabprofiler of the post-writing of learners from NnDG**

| Participants | Word Count | 1k+2k % in Text | AWL | | | | |
|---|---|---|---|---|---|---|---|
| | | | Families | Types | Tokens | Percent/AllWards | Percent/150Wds |
| 315 | 222 | 98.64% | 0 | 0 | 0 | 0.00% | 0.00% |
| 320 | 145 | 93.01% | 0 | 0 | 0 | 0.00% | 0.00% |

### *5.4.2.3. The relations between AWL abilities and the lexical choice in free writing:*

As stated earlier, with such a small population, the data would be insufficient to perform statistical conclusions and, therefore, the entire data of each participant was reported earlier. However, it was thought that it would still be interesting to look at any expected effect of the treatment or any variation of improvement based on groups, since previous analyses of the data regarding the three multiple measurements in this chapter confirmed the variation in terms of improvement depending on groups (i.e. DG vs. NnDG). In other words, it is believed that the significance due to the group effect addressed earlier in the chapter which investigated the whole population (n = 86) taking three different measures, could similarly be found with this small population size (n=8). Any variation in performance found between those learners based on group effect in the three measures could give some indication about the role of the effect in the variation in the fourth measure (i.e. lexical profile). If the variation was noticed in this small sample size, then the variation discussed earlier regarding the AWL lexical profile in writing (section 5.4.2.2) is likely to be true, regardless of the limitations discussed. Of course, only descriptive statistics and graph illustrations were shown, due to the insufficient data that can be used to perform statistical representations.

The data of all measures including the writing analysis discussed in the previous section from the 4 students of DG and the 4 students of NnDG were included in this descriptive statistics. It was decided here to include the 2 students from NnDG who had different writing from their counterparts in the comparisons to explore their overall abilities in all measures, and disregard the variation in the writing task. The descriptive results of all participants in both groups are summarized and graphed below. A note about the students who had different writing tasks is also indicated.

**Table 5.31 Descriptive reports of learners' test scores and their AWL lexical frequencies:**
**A) Scores from learners of DG:**

| Ss ID | VLT or REC | | XKLEX | | CATSS | | AWL | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Pre | post | pre | post | Family | | Percentage | |
| 210 | 80 | 83 | 70.0 | 63.0 | 30 | 56.6 | 10 | 7 | 3.90% | 2.35% |
| 222 | 86 | 83 | 51.0 | 51.0 | 36.6 | 56.6 | 10 | 4 | 4.61% | 2.29% |
| 232 | 53 | 43 | 44.0 | 37.0 | 3.3 | 6.6 | 1 | 1 | 0.96% | 2.40% |
| 234 | 73 | 86 | 20.0 | 52.0 | 13.3 | 30 | 5 | 3 | 6.63% | 1.97% |

**B) Scores from learners of NnDG:**

| Ss ID | VLT or REC | | XKLEX | | CATSS | | AWL | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Family | | Percentage | |
| 315 | 63.3 | 83.3 | 44.0 | 34.0 | 0 | 16.6 | 6 | 0 | 3.63% | 0.00% |
| 320 | 73.3 | 70 | 36.0 | 38.0 | 0 | 13.3 | 4 | 0 | 3.79% | 0.00% |

**B) = Scores from the two NnDG participants who had different second writing task:**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 301 | 60 | 53 | 27.0 | 33.0 | 10 | 6.6 | 10 | 3 | 4.81% | 1.71% |
| 306 | 30 | 13.3 | 6.0 | 20.0 | 6.6 | 6.6 | 2[a] | 3 | 5.93% | 1.81% |

*a. 2 word families but 8 tokens.*

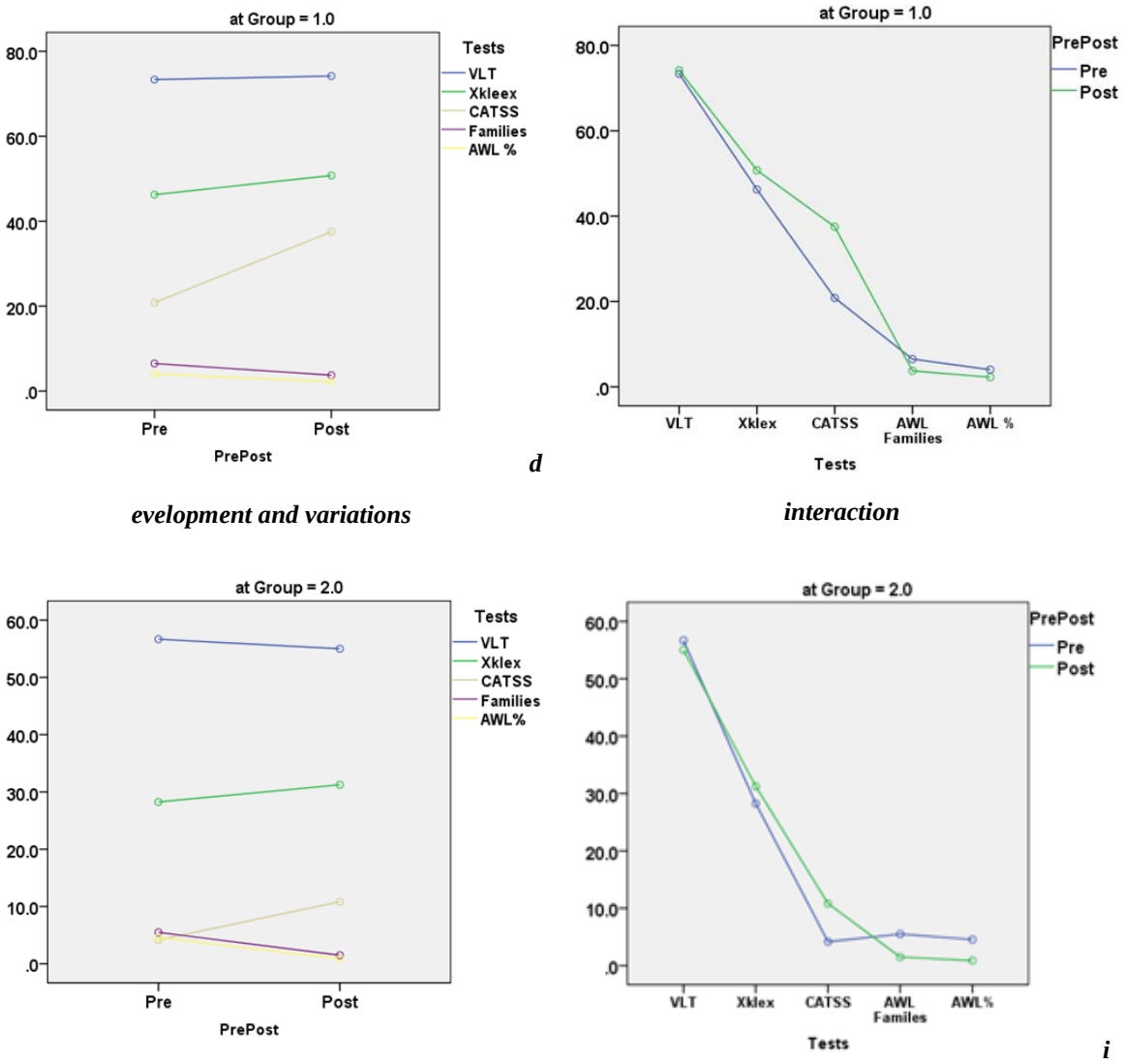



**Figure 5-8: Pre-test Post-tests and lexical profile graphical interactions of each group n=4**

The tables do not show any clear-cut patterns regarding variations in improvement between the two groups, nor interactions between measures and the AWL lexical profile. The two patterns that emerge are that learners from the DG scored more highly among the measures than learners from the NnDG, and that there might be a linear interaction between the controlled ability to produce the AWL (CATSS results) and the free AWL lexical profile in writing. This pattern matches with the findings in Chapter 4 (see the

discussed in section 4.5.5.3). Of course, it was obvious from the targeted population of this study that learners from the DG group had higher initial scores than NnDG learners.

The graphs may give some idea about a possible variations between learners based on groups in the two productive measurers (CATSS and AWL %). Both of the groups increased their scores in CATSS; however the scores of learners exposed to extra input (treatment) increased more sharply. Regarding the percentage of AWL in free writing, both have decreased (as justified in the previous section), but the percentages for learners from the NnDG decreased more sharply. The figures in the tables give some idea about the levels in all measures by each student in both groups.

## 5.5. Summary of the chapter

This study explores the learning outcomes resulting from 'two different methods of learning the academic words list', 'by advanced learners', 'using multiple tests to assess vocabulary mastery'. Many studies reviewed in this thesis have discussed the issues of incidental learning of vocabulary vs. direct or enhanced teaching of vocabulary to either beginners or advanced EFL learners. A number of studies that involved advanced students were considered and several arguments were made about the effective methods of teaching vocabulary to advanced learners (see sections 2.3.2.1 and 2.3.2.2). One of the main points addressed was that the common perception among some educators regarding advanced students is that they are capable of acquiring more language input incidentally, and therefore, less direct instruction is applied (more discussion on this can be found in in 1.4.). In addition, due to the incremental nature of vocabulary knowledge discussed earlier in the thesis, multiple measures were used to assess the different mastery levels of the AWL. Furthermore, there is some vagueness regarding the most effective way to

teach vocabulary to advanced students. i.e. since advanced EFL learners are considered capable of grasping language input incidentally, especially vocabulary, this study investigated how well they in fact know the words. Results from the *first study* confirmed that many advanced learners showed higher levels of receptive knowledge of the AWL, but lower abilities regarding the productive ability of the same word list. This study also explored variations in improvement amongst the learners based on the method of enhancement of the AWL (i.e. variations according to group) considering the different levels of AWL mastery.

The study applied the same multiple measurements used in the *first study*: the AWL part of VLT for the receptive knowledge, CATSS for the controlled AWL productive knowledge, analysing the use of AWL freely in writing, and the Xklex to estimate the general size. Two similar groups of Saudi graduates studying on an intensive English foundation programme participated in the study. One of the groups (Direct Group DG) was exposed to 'normal' academic English study in the programme as well as outside direct enhancement that targeted the AWL, and the other group (Non-direct NnDG) was exposed to normal academic English instruction only.

The results generally showed significant variation in terms of improvement in favour of the treatment group. Firstly, regarding the receptive knowledge, the initial comparison analysis was not able to capture any variation between the groups simply because learners mostly reached the maximum score in the pre-test and this did not increase. The mean differences between pre- and post-test in both groups were not significant initially. A more robust test was performed which confirmed the significance of improvement with DG learners only. Regarding CATSS, based on the first study, the increase was far more

obvious than in the VLT. However, interestingly, the knowledge of learners from the treatment group DG increased more significantly than that of the NnDG group, even though they had similar pre-scores.

Regarding general vocabulary size, learners were not expected to vary in this measure since the treatment might not directly contribute to this knowledge. However, unexpectedly, learners' knowledge did in fact decrease in the post-test. It was suspected that the cause of this unexpected decrease was the adjustment made to the original construction of the test. A follow up study was conducted to investigate this but did not confirm this suspicion.

Lastly, regarding the variation in the AWL lexical profile in writing, there was no significance of variation based on group due to some limitations discussed in section 5.4.2.1. However, some indirect analysis could conclude that some variation may have occurred between learners, based on groups.

The results from this study are further discussed in Chapter 7.

# 6. Chapter Six

# Learners' and Teachers' feedback on learning the academic word list

## 6.1. Introduction and aim of the study

As argued previously in section 2.2.4., recently there has been a great deal of language learning literature that explores feedback from learners in order to study their beliefs and experience about learning (see section 2.2.4). Their opinions are very useful as they evaluate the learning process, help in identifying learners' needs or areas of difficulty, and make educators aware of their beliefs or theories about what influences students' learning. Moreover, feedback surveys that occur after experiments and empirical studies help greatly in analysing the data and help the researcher develop a clearer understanding of results. In other words, as argued previously in section 3.4, this study was treated as an independent study as the feedback data that it involved addresses a number of issues related to teaching vocabulary to advanced learners. However, since the participants of this study are all of the learners involved in the previous two studies, some of the data of this survey would be usable and useful in clarifying or justifying any unexpected results regarding some individual learners within whole population of study one and two.

Although there are many studies that have explored students' behaviour and strategy use in learning vocabulary, it has been noted that there are fewer publications that have solely discussed beliefs and the personal theories of learners which may influence their vocabulary learning, apart from a few studies that used *BALLI*[6] for general language use (see section 2.2.4.)

There is little in the literature that has specifically looked into learners' beliefs about vocabulary. With advanced learners, research into beliefs about vocabulary learning

[6] *Beliefs About Language Learning Inventory*

could be of even more benefit since the general view is that capable students are given less focus on lexical instruction (Alothman, 2011: see 1.4.).

With respect to the discussion above, this study aims to address the following research question:

> **RQ 3-** What are learners' and teachers' perceptions and beliefs about vocabulary teaching at this level of proficiency.

In order to respond to this research question and to the research hypothesis discussed in section 3.3, this study attempts to:

1- Investigate advanced learners' views on vocabulary learning in general and academic vocabulary in particular.

2- Study the behaviour of learners regarding learning academic words on advanced and intensive English programmes.

3- Address the learning capabilities of advanced learners on presessional and intensive English courses by giving them the chance to express their needs/or declare their ability of performance of academic words.

4- Explore some of the learning strategies that advanced learners use to acquire the academic words in advanced courses.

5- Investigate whether the above points vary with respect to learners' language proficiency or their academic major after the courses.

*The purpose of teachers' survey:*

Teachers of presessional courses and PY programme were surveyed to explore their views on the learning academic words by advanced learners. The survey of teachers

would be useful in describing actual classroom practices regarding teaching the AWL on advanced courses. It would show the beliefs of teachers regarding the issue, and would be an opportunity to describe how motivated they felt their students were during the programme of study. Regarding the second study (Chapter *5*), teachers of the active classes (treatment or direct group DG) might give useful insights into the direct method of teaching the AWL used by them during the programme; how feasible the amount of words to be taught was, and their thoughts about the efficiency of the method.

## 6.2. Method

### 6.2.1. Procedure.

After the questionnaires were designed and distributed to all participants directly after the posttest was administered. As outlined in section 4.3, all participants were aware of their rights regarding confidentiality and access to their data. Although writing their names on the questionnaire sheets was optional, they were encouraged to do this for the purpose of this study only (i.e. to correlate data between their multiple tests and their feedback). They were assured that their personal data would not be accessed by anybody except for research purposes only, and that neither the administration nor their teachers would have access to the data.

Concerning the teachers, they were given open-ended questionnaires after the students were tested; this was done during their office hours. Their names were not as important for the study as their students'; however, they were asked to write the groups/classes they taught in order to have an idea about any possible factor related to the groups in previous studies (*First* and *Second* studies). Teachers were given extra space on the questionnaire sheets to write more remarks if they wished.

### 6.2.2. Participants

The participants of this study are all students involved in the other studies in Chapter 4 and Chapter 5, plus their teachers. However, since there were multiple tests performed in the previous studies, and each participant had to fill in a number of answer and information sheets, there were some student papers missing. It was decided in the first two studies to exclude any students who missed either one of the pre-or post-tests, or missed one of the multiple tests (VLT, CATSS and XKlex). In this study, however, it was believed that the surveys would not be affected if a learner missed some of the tests. Table 6.1 below presents a summary of the final number of the participants and teachers with more details about all respondents:

**Table 6.1 Overview of participants in this study:**

| | Major | | | | PG, or UnderG | |
|---|---|---|---|---|---|---|
| **Students' Groups [a]** | **Humanities** | **Medicine** | **Science** | **Pre-study** | **PG** | **UG** |
| **SUni 1** | - | - | - | 48 | 42 | 6 |
| **RUni 2** | - | - | - | 43 | 41 | 2 |
| **SA-DG** | 33 | 8 | - | - | 41 | - |
| **SA-NnDG** | 27 | 11 | 1 | - | 39 | - |
| **Total 171** | 60 | 19 | 1 | 91 | 163 | 8 |

[a] as noted before, study one: two universities involved (SUni1, RUni2), study two: two Saudi groups at one university but were involved two groups, and study three included all participants as explained below.

In the UK-based investigation (SUni1 and RUni2), the participants were of different nationalities. There were students from China, Saudi Arabia, Spain, and Libya. In the investigation that took place at the PY- SA, all participants were male Saudi students. Teachers who participated in the survey were all native speakers of English, but of different nationalities (UK, US and Canada).

### 6.2.3. The instruments.

*Students*: students were given a 17-item questionnaire that was developed to gather some information from them that might be useful for the study and to identify, generally, their learning needs and views about learning academic words. It also looked to find out their common learning strategies, and to access their thoughts on the best methods of learning. This questionnaire was adopted and modified from an original questionnaire that was designed and used in (Alothman, 2009) (see Appendix J). The current questionnaire was proven to be reliable as shown below (see section 6.4.1.1). The questionnaire was designed to be very focused and comprehensive and, therefore had few survey items. Due to time and space logistics, it was believed that longer questionnaires would interrupt more of learners' school time and learning schedules. Since all students would have to take multiple tests, it would have been too much for them to fill in all papers with comfort and ease. If they had, the performance in all measures could have been affected. Lastly, it was thought that comprehensive and focused survey items would be more attractive and would encourage more students to participate in the study. This would also encourage the schools to agree to the multifaceted study taking place.

Although the questionnaire was kept short, it was designed to meet at least the minimum requirement items that address the research question of this chapter. The questionnaire was originally 16 items. However, as noted at the end of Chapter 4, this was increased to 17 in the *First Study* after amending one item and breaking it into two separate statements. The item amended was item 11 and it was broken into 11 and 12. It was thought that the original item might have been too complex and may be holding 2 answers. The original statement and amendments are shown below:

**The original item in the preliminary questionnaire:**

| |
|---|
| 11. I have been taught some academic vocabulary and I think it is enough for my further studies. |

**After the amendments**

| |
|---|
| 11. I have been taught some academic vocabulary in class. |
| 12. The academic vocabulary I now know is enough for my further studies. |

The items were also modified to accommodate the targeted participants, depending on context and location. The items were divided into three parts; the first part (items 1-4) asked respondents for some information about their personal background, their class or group within the school, some information about proficiency, study major, and undergraduate/post graduate status. Part Two (5 to 14- or to 15) asked participants to give their views about statements using a 5-point Likert scale ranging from *strongly agree* to *strongly disagree* (see the following example and Appendix J).

| | | | | | |
|---|---|---|---|---|---|
| 9. The academic vocabulary is important for my language study. | **Disagree** | | | | **Agree** |
| | **1** | **2** | **3** | **4** | **5** |

In this part of the questionnaire they were asked to give their views regarding the following:

**Figure 6-1: Categories for the questionnaire items**

| | |
|---|---|
| 1- Describe their learning of the AWL in EAP courses and give their opinions on the current practices. | Items 5, 6, 7, 8, 11, 14 and 15 |
| 2- State their opinions on the current practices. | 12 and 13 |
| 3- Declare their general beliefs about the academic words. | 9 and 10 |

Part Three had two items; one asked a yes/no question about use of vocabulary books plus an additional blank space if the answer was yes to name the book. The other item was a statement that asked them to describe their most common practice during their programmes regarding learning the AWL.

Because the participants were advanced students, it was thought that there was no need for the items to be translated, even in the context where all participants were Arabic speakers.

*Teachers:* Teachers were given a 7-item semi open-ended survey that asked them to evaluate the learning of academic vocabulary on their English courses, and state some of their views about teaching the AWL. It also gave them the opportunity to describe how motivated the students were to learn the AWL, and how many of the words on the list they thought could feasibly be taught on their curriculum.

The 7 items were divided into two parts; the first part (items 1-3) involved some statements that asked teachers to state their views about each given statement. They ticked the response that best described their view on the statement given. Here is their agreement/disagreement scale;

| | | | | |
|---|---|---|---|---|
| o Strongly agree | o Agree | o No opinion | o Disagree | o Strongly Disagree |

These items generally asked teachers about teaching the AWL on their programmes and their beliefs about the classroom practices that they experienced. Items 4 to 7 were open-ended questions and included some statements rating actual classroom practices. The open-ended questions asked teachers to describe how motivated their learners were to learn the AWL, what had been covered in the programmes from the list, and what they thought was feasible for teaching in a term. Items 4 and 7 asked teachers to describe or state which situation regarding teaching the AWL had commonly occurred in their classes, and to state from their own experience how learning the AWL impacted on learners' skills.

## 6.3. Data analysis

*Students' questionnaires:* the questionnaire items and statements were given numbers, or shorter names, before they were entered in *Microsoft Excel* tables. Learner' background such as group or school, level of proficiency, major and level of study were transformed into numerical codes to be run in *SPSS*. Sometimes, an extra column was created to include infrequent responses such as the score of a recent proficiency test or even the type of test if it was different from the tests mentioned in the questionnaire sheet; i.e. some schools have local tests similar to IELTS or TOEFL. Responses to the learning of academic vocabulary (i.e. items 5 to15 see 6.2.3) were scored according to a Likert 5-point scale: (1) for *totally disagree,* (2) *disagree,* (3) neutral, (4) *agree*, and (5) *totally agree*.

It useful to note that due to the difference in the number of items between the preliminary questionnaire (Chapter 4 participants), and the second questionnaire (Chapter 5 participants), discussed in the earlier section, an extra blank column was added when coding the preliminary questionnaire to guarantee equality of the number of items and consistency of organization of all questionnaires. The values of *item 11* of the preliminary questionnaire were repeated in the blank column.

All questionnaire items were checked for reliability using SPSS as explained in section see 3.4.3.

The statistical SPSS package was employed to gain a descriptive analysis of the responses. Multivariate analysis of variance *MANOVA* and covariant *MANCOVA* were performed to identify any significance among means of response rates of participants in their respective groups.

To serve the purpose and the design of the study, different types of relationship analysis were employed:

- 1. The interactions between learners' background (specially school or group) and their responses about beliefs regarding the AWL, and describing the actual practice in the programmes.
- 2. The interactions between their responses and their proficiency level.
- 3. The interactions between their responses and their performance in the tests of the studies in Chapters 4 and 5.

*Teachers' open-ended questionnaires:* similar to students' questionnaire, the items of this questionnaire were given numbers or shorter names before entry into *Excel.* Their responses to items (questions 1-3), were transformed into numerical codes to represent their views in the 5-point scale: (1) for *totally disagree,* (2) *disagree,* (3) *no opinion*, (4) *agree,* and (5) *totally agree*. The same procedure was followed with item 5, which asks teachers about learners' motivation, and item 7 which asks for their evaluation of the impact of the AWL in different linguistics skills. The scale was as follows: (1) *highly,* (2) *somewhat*, (3) *not sure*, (4) *little*, (5) *not at all*. The statistical methods *ANOVA* and *Correlation* were used in order to describe the analysis of the responses. Question 4 has five items to which no scale was applied, and allowed choice of more than one option. Descriptive analysis is applied here. The comments about coverage of the AWL in classes in question 6 were analyzed so that they might be displayed in tables according to the research questions.

The following correlations and descriptive analysis were performed to serve the purpose and the design of the study

- 1. Descriptive analysis of teachers' responses regarding beliefs and practices about the AWL.
- 2. Correlation between teachers' groups, and their learners' scores in the tests of all the tests used in the studies in sections 3.1. and 3.2.
- 3. Responses about beliefs regarding the AWL and their descriptions of the actual practices in the classes.

## 6.4. Results and discussion

The research question and hypothesis involve exploring:

- Learners' beliefs as well as needs regarding the AWL on intensive courses.
- Learners and teachers' feedback about the way vocabulary is taught on the courses they are involved in.

The question and hypothesis impose deeper queries that consider learners' background differences, which may affect their responses.

The results show some general quantitative statistics as well as some specific qualitative data. First, a general statistical summary of all statements is presented then the main factors are considered and addressed in turn below.

### 6.4.1. Descriptive data of all responses of learners in general and in groups.

#### *6.4.1.1. Participants' background*

As explained in the general procedures of the thesis, the participants of this study are all learners who participated in the *First* and *Second studies* (see also Table 6.1. above). The background details plus the *grouping* factors are considered in the following analysis. A

summary of all learners' background data from the questionnaire (items 1, 2 and 4) is shown below in Table 6.2. and their details regarding English proficiency is in Table 6.3.

**Table 6.2 Participants' background**

| | Years of EFL study | | | | | Major | | | | PG,or UnderG | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | 0-5 | 6-7 | 8-9 | 10-11 | Over | Human | Med | Science | N/A | PG | UG |
| SUni 1 | 1 | 17 | 12 | 4 | 16 | - | - | - | 50 | 42 | 6 |
| RUni 2 | 3 | 6 | 3 | 5 | 27 | - | - | - | 44 | 41 | 2 |
| SA-DG | 1 | 6 | 6 | 5 | 23 | 33 | 8 | - | - | 41 | - |
| SA-NnDG | 0 | 6 | 15 | 5 | 13 | 27 | 11 | 1 | - | 39 | - |
| Total:174 | 4 | 35 | 36 | 19 | 79 | 60 | 19 | 1 | 94 | 163 | 8[1] |

[1] Three participants declared none.

**Table 6.3 Participants' proficiency tests**

| | Taken proficiency test | | IELTS score or equal | | | | |
|---|---|---|---|---|---|---|---|
| Group | None | IELTS-TOEFL-TEEP-IBT | 4.5 | 5.5 | 6 | 6.5 | 7 |
| SUni 1 | 13 | 37 | - | 26 | 6 | 2 | 3 |
| RUni 2 | 12 | 32 | - | 17 | 6 | 1 | 7 |
| SA-DG | 25 | 16 | 1 | 6 | 3 | 3 | 3 |
| SA-NnDG | 30 | 9 | - | 6 | 3 | - | - |
| Total:174 | 80 | 69 | 1 | 55 | 18 | 6 | 13 |

It is predictable that there would be a large variation between the participants of this study, since the first two groups are from UK presessional programmes, and the other two groups are Saudi PY foundation academic students. Added to this, the majority of the UK participants are postgraduates, and all Saudis are undergraduates. However, as has been argued before, there are some research gaps concerning learners' beliefs and their statements regarding learning and teaching the academic vocabulary on their intensive advanced English programmes (see further discussion in sections 2.2.4. and 6.1.). With regard to the variations between the UK and Saudi groups, the data however, shows some similarities between those advanced EFL learners concerning the estimate of their

proficiency level. The available data shows that the level of the majority of participants in all groups ranged between IELTS 5.5 and 6.0.

The analysis below presents general descriptive statistics about a number of advanced EFL students' perceptions regarding the AWL learning, and considers all factors that may affect learners' views.

In terms of participants' background data, the obvious difference is the postgraduate/undergraduate status of the UK participants and Saudi participants. Participants' statements about *years of study* seemed similar in terms of UK presessional vs. Saudi PY programmes. Regarding major of study, the data was only available with the Saudi participants. Within the two Saudi groups, there was *no huge* difference in this category among learners, i.e. they had the same distribution within the majors of study. Any inequality in this factor, if it was the case, might be very significant regarding influencing the rates or motivation of learners regarding learning the AWL (see also the discussion in section 5.2.1.). As mentioned, a summary of the general views of all learners is shown below, and then a discussion of all possible factors will follow.

*Reliability and normality*

The questionnaire items 5 to 17 proved to have acceptable (considering the small number of items) overall reliability using *Cronbach's Alpha* (.680). Regarding normality, the items proved to be normally distributed using the two-sample Kolmogorov-Smirnov test, except for items 11 and 12. Obviously, the reason for the non-normal distribution with those two items could be the amendment conducted after *the first study* as discussed before in section 6.2.3. i.e. item 11 was broken into 11 and 12 only with participants in the second study.

*6.4.1.2. Descriptive statistics from all respondents:*

A direct answer to the research questions raised above in this chapter is that advanced learners who took part in this study declared their need to learn academic words in their advanced and intensive programmes, and that some learners appreciated the enhancement of the AWL in their courses where it was offered to them. Learners believe that the AWL is important for their English studies (rated 4.2 out of 5), and for their academic studies (rated 4.16 out of 5). Learners mostly think that the words of the AWL they learned by the end of their programmes were not enough for their graduate/postgraduate studies. Learners who confirmed learning the AWL on the courses in fact wished they had more. Not all advanced learners reported learning the AWL during the course, nor being able to use the words in speaking and writing. 44 % of learners confirmed using a textbook to study vocabulary, and almost half of them reported learning the AWL independently as the table shows. Table 6.4 is a summary of their beliefs, statements of learning needs and feedback regarding learning the AWL.

**Table 6.4 Descriptive statistics of all respondents**

| Statements | Rate | | |
|---|---|---|---|
| | Disagree 1 2 3 4 5 Agree | | |
| | Mean rate | SD | N |
| 5. I learned a lot of academic vocabulary in this course. | 3.47 | 1.146 | 174 |
| 6. I learned a lot AWL and could write and speak them | 3.33 | .987 | 174 |
| 7. I'd specific focus on the AWL during the course. | 3.38 | 1.209 | 174 |
| 8. The AWL was enhanced within course language skills | 3.45 | 1.193 | 173 |
| 9. The AWL is important for my language study. | 4.20 | 1.196 | 174 |
| 10. The AWL is important for my further studies. | 4.16 | 1.279 | 173 |
| 11. I have been taught some academic vocabulary in class. | 3.06 | 1.494 | 174 |
| 12. The AWL I now know is enough for my PG studies. | 2.59 | 1.330 | 174 |
| 13. I learned AWL, but I wished I had more during course | 3.60 | 1.225 | 174 |
| 14. I studied vocabulary outside class with teacher's support. | 2.93 | 1.105 | 174 |
| 15. I studied vocabulary outside class on my own. | 3.11 | 1.096 | 174 |

Participants' responses regarding item 16 about the use of a certain textbook to learn the academic words during their programmes:

| | Yes % | N |
|---|---|---|
| Did you use any textbook which focuses on AWL during the course: | 44 | 174 |

The responses to item 17 to describe learning the general AWL learning during this term:

How would you describe your vocabulary learning during this term? (more than one option is possible):

| | % | SD | N |
|---|---|---|---|
| A) In class aided by teacher | %72 | .453 | 172 |
| B) Homework set by teacher. | %42 | .506 | 172 |
| C) Self vocabulary learning. | %49 | .535 | 172 |
| D) None. | %02 | .152 | 170 |

### *6.4.1.3. Further investigation of variables:*

Further analyses were performed to investigate possible effects that could have an influence on learners' beliefs and feedback about learning the academic words. The main Potential factors that were believed to cause the difference among respondents were: groups, UK vs. Saudi learners, level of study, major of study, level or test of proficiency and years of English EFL study. An investigation was also performed to explore the effect of *group* within the Saudi participants (i.e. DG vs. NnDG) since it involved two different learning groups. Table 6.5 below is a summary of learners' mean rates in each group. The small letters beside the means indicate a *significant difference*. Each letter represents a variable as discussed shortly:

**Table 6.5 Descriptive statistics of responses in their respective groups**

| Statements | Disagree 1 | 2 | 3 | 4 | Agree 5 |
|---|---|---|---|---|---|
| | SUni1 | RUni2 | SA-DG | | SA-NnDG |
| 5. I learned a lot of academic vocabulary in this course. | 2.84 | 3.77 | 4.07 a c * | | 3.28 |
| 6. I learned a lot AWL and could write and speak them | 2.82 b * | 3.36 | 3.9 c | | 3.36 |
| 7. I'd specific focus on the AWL during the course. | 3 | 3.55 | 3.93 a c * | | 3.1 |
| 8. The AWL was enhanced within course language skills | 2.82 a * | 3.82 | 3.88 | | 3.39 |
| 9. The AWL is important for my language study. | 3.96 | 4.16 | 4.76 c * | | 3.95 |
| 10. The AWL is important for my further studies. | 3.8 | 4.27 | 4.61 c | | 4.03 |
| 11. I have been taught some academic vocabulary in class. | 2.24 b | 2.05 b | 4.44 a c | | 3.82 |
| 12. The AWL I now know is enough for my PG studies. | 2.24 b | 2.05 b | 3.17 | | 3.05 |
| 13. I learned AWL, but I wished I had more during course | 3.18 | 3.77 | 3.61 | | 3.95 |
| 14. I studied vocabulary outside class with teacher's support. | 2.66 | 3.07 | 3.15 | | 2.9 |
| 15. I studied vocabulary outside class on my own. | 3.22 | 2.98 | 3.02 | | 3.23 |

a, b and c are the variables that were found to have significant effects: a) Group, b) UK vs. Saudi participants, and c) Saudi DG vs. NnDG, see the discussion below.

P >.05, *P > .005,

Item number 16 of learners within groups.

| | Disagree 1 | 2 | 3 | 4 | Agree 5 |
|---|---|---|---|---|---|
| | SUni1 | RUni2 | SA-DG | | SA-NnDG |
| Did you use any textbook which focuses on AWL during the course: **%Yes** | 56 | 48 | 41 | | 28 |

Item number 17:

17. How would you describe your vocabulary learning during this term? (more than one option is possible) **(%)**

| | Disagree 1 | 2 | 3 | 4 | Agree 5 |
|---|---|---|---|---|---|
| | SUni1 | RUni2 | SA-DG | | SA-NnDG |
| A) In class aided by teacher | 82 | 56 | 85 c | | .62 |
| B) Homework set by teacher. | 53 | 79 | 17 b * | | 13 b * |
| C) Self-study vocabulary learning. | 67 | 37 | 39 | | 49 |
| D) None. | 0 | 0 | 03 | | 08 |

A number of *MANOVA* multivariate analysis of variance and tests were performed to check the interactions and effect of variants. Three main variants showed significant differences (effects) among group rates and were thought to influence views. The main factors were: *a)* Group, *b)* UK vs. Saudi participants, and *c)* Saudi DG vs. NnDG. The remaining variables showed either insignificant variations or showed similar interaction to the main effects above. The output tables of the *MANOVAs* of the all variables are shown in Table 6.6 below, and the further tests of between-items effects of the main variable (Group, UK vs. Saudi participants and Saudi DG vs. NnDG) are shown in Tables 6.7, 6.8 and 6.9 at the end of the this section, and the remaining variables are shown in Appendix K.

**Table 6.6 Multivariate Tests of effects**

| Effect | *F* | Hypothesis df | Error df | Sig |
|---|---|---|---|---|
| Group | 6.642 | 48.000 | 443.957 | .000 a |
| UK vs SA groups | 7.547[a] | 16.000 | 150.000 | .000 b |
| Saudi DG vs. NnDG | 2.638[a] | 16.000 | 60.000 | .003 c |
| Test score | 1.043 | 64.000 | 268.484 | .399 |
| EFL year study | .980 | 64.000 | 581.670 | .522 |
| Level of study | 4.081 | 48.000 | 443.957 | .000 1 |
| Major of study | .902[a] | 32.000 | 118.000 | .621 |

a, b and c are the variables that were found to have significant effects,

1. *'Level of study'* was found to overlap with '*UK vs SA'* and explained.

A) *Variations according to groups;* the rates of item 5 ''learning the AWL during the course'' by Saudi DG learners was significantly higher than the other three groups at p > .005. The same significant difference was found in this particular group in items 7 ''had specific focus on the AWL during the course'' and at p >= .005 in item 11 ''have been taught some academic words in class''. A smaller percentage of students in the SUni1 group report than any of the other groups for Item 8 ''agreeing that the AWL was enhanced within the course language skills''. This is expected considering this group's

responses regarding items 5, 6, 7, and 11. i.e. they all indicate less enhancement of the AWL in their course.

B) *Variations between UK and Saudi participants*: first, it is important to note that this variation factor is in fact in line with another factor which concerns 'level of study', i.e. the majority of UK learners were postgraduates and all Saudis were undergraduates. Therefore, they are likely to overlap. It is also important to note that the variation between responses should have more factors due to the vast variations between the two learning environments as addressed above in section 6.4.1.1. In terms of significance in the variations of learners' rates according to group location, the *MANOVA* showed significance differences among respondents in item 6 ''being able to speak and write the AWL''. SUni1 learners reported that they were the least able to speak and write the AWL. The rate of items 11 and 12 was also significantly lower among UK learners, but that could be due to the amendment and editing that affected these two items as discussed before. I.e. the two statements were originally in one item in the questionnaire sheets given to the UK participants. This is what could have affected the low rates regarding these two items. Lastly, with regard to the second option in item 17, the UK learners reporting their ''studying of the AWL as homework set by teacher'' were significantly more than the Saudi learners.

C) *Variations among Saudi learners DG vs. NnDG groups*: variation in responses among the Saudi learners depending on their learning group was expected since they experienced two different academic vocabulary learning methods (more information on this is in Chapter 5). As discussed in the previous section, the learners in both groups shared almost the same background regarding study and proficiency. The responses of the

learners of the treatment group DG had significantly higher agreement rates with statements than NnDG learners in the following items: item 5 ‘‘Learned a lot of the AWL in this course’’, item 6 ‘‘learned the AWL and could write and speak them’’, item 7 ‘‘had specific focus on the AWL’’, and item 11 ‘‘have been taught the AWL in class’’. Also, regarding item 17 which asked participants to describe the learning of the AWL during the term by each learner, the number of learners of DG was significantly more than the NnDG learners. These variations were expected, and they are in line with the results found in Chapter 5. To illustrate briefly, the responses explain and describe the impact and effect of the treatment regarding academic vocabulary enhancement explained in the *Second* study. i.e. since the DG group of learners were exposed to the extra treatment of the AWL, they reported higher in this statement than learners of the other groups. However, there was a significance difference between the learners regarding their beliefs and thoughts about the statements in items 9 and 10. The DG learners’ agreement about the importance of the AWL was significantly higher than that of NnDG learners. This could mean that the significant difference among learners’ outcomes according to active or treatment group DG vs. the passive group NnDG that was addressed in Chapter 5 was in fact due to the motivation of DG learners. However, the effect of the treatment factor was expected and therefore is more obvious. In addition, the difference in the beliefs could be in fact an effect not a cause. i.e. the fact that learners from the DG group were the only group receiving this extra treatment could influence their responses to this item.

Apart from the above variables, as shown in Table 6.6 above, all the respondents seemed to have similar views regarding the statements. There was no variation among learners in terms of major of study, number of years of EFL education, or proficiency. As explained,

the level of study (graduate/postgraduate) presented similar interaction results to the factor *UK learner vs. Saudi learners*. Below are Tables 6.7, 6.8 and 6.9 showing further tests of *between-items effects* of the main variable (Group, UK vs. Saudi participants and Saudi DG vs. NnDG). The remaining variables as mentioned earlier are shown in Appendix K. These tables further explain Table 6.5 above.

**Table 6.7 Tests of Between-Subjects Effects: Group**

| | | Value Label | N |
|---|---|---|---|
| **Group** | **1** | **SUNI1** | **49** |
| | **2** | **RUni2** | **42** |
| | **3** | **Saudi DG** | **40** |
| | **4** | **Saudi NnDG** | **37** |

| Dependent Variable | Df | M-Square | F | Sig. |
|---|---|---|---|---|
| **Rates**: | | | | |
| Learned AWL | 3 | 13.816 | 12.984 | .000 |
| Learn And Can Write Speak AWL | 3 | 8.702 | 10.226 | .000 |
| Course Focused On AWL | 3 | 7.268 | 5.412 | .001 |
| AWL was Enhanced within Skills | 3 | 11.075 | 8.805 | .000 |
| AWL is important for EFL | 3 | 4.977 | 3.971 | .009 |
| AWL is important for PostGrad | 3 | 4.711 | 3.136 | .027 |
| I've Been Taught AWL in Class | 3 | 57.839 | 47.666 | .000 |
| AWL I learned Was Enough | 3 | 13.223 | 8.427 | .000 |
| I Wished I Had More AWL | 3 | 5.532 | 4.000 | .009 |
| I Had AWL out Class Set By Teacher | 3 | 1.996 | 1.641 | .182 |
| I learned AWL out Class By My Self | 3 | .818 | .672 | .570 |
| **Questions**: | | | | |
| I Used Text Books to study AWL | 3 | .546 | 2.240 | .086 |
| I Had AWL in Class | 3 | .910 | 4.732 | .003 |
| Had Homework By Teacher | 3 | 4.044 | 21.606 | .000 |
| I Self Learned the AWL | 3 | .893 | 3.231 | .024 |
| No AWL study This Term | 3 | .058 | 2.534 | .059 |
| Error | 164 | | | |

**Table 6.8 Tests of Between-Subjects Effects: UK vs. SA**

| | | Value Label | N |
|---|---|---|---|
| UK vs. SA | 1.00 | UK | 91 |
| | 2.00 | SA | 77 |

| Source | Dependent Variable | Df | M- Square | F | Sig. |
|---|---|---|---|---|---|
| **Rates**: | | | | | |
| Learned AWL | | 1 | .102 | .081 | .776 |
| Learn And Can Write Speak AWL | | 1 | 1.595 | 1.733 | .190 |
| Course Focused On AWL | | 1 | .957 | .659 | .418 |
| AWL was Enhanced within Skills | | 1 | 1.139 | .817 | .367 |
| AWL is important for EFL | | 1 | 4.516 | 3.471 | .064 |
| AWL is important for PostGrad | | 1 | .414 | .265 | .607 |
| I've Been Taught AWL in Class | | 1 | 66.727 | 54.976 | .000 |
| AWL I learned Was Enough | | 1 | 16.142 | 10.347 | .002 |
| I Wished I Had More AWL | | 1 | 4.901 | 3.556 | .061 |
| I Had AWL out Class Set By Teacher | | 1 | .012 | .010 | .922 |
| I learned AWL out Class By My Self | | 1 | .213 | .174 | .677 |
| **Questions**: | | | | | |
| I Used Text Books to study AWL | | 1 | .028 | .117 | .733 |
| I Had AWL in Class | | 1 | 2.519 | 13.167 | .000 |
| Had Homework By Teacher | | 1 | 4.885 | 25.623 | .000 |
| I Self Learned the AWL | | 1 | .092 | .325 | .570 |
| No AWL study This Term | | 1 | .000 | .000 | .999 |
| Error | | 164 | | | |

**Table 6.9 Tests of Between-Subjects Effects: DG vs. NnDG**

| | | Value Label | N |
|---|---|---|---|
| Group | 3 | Saudi DG | 40 |
| | 4 | Saudi NnDG | 37 |

| Source | Dependent Variable | Df | M- Square | F | Sig. |
|---|---|---|---|---|---|
| Rates: | | | | | |
| Learned AWL | | 1 | 12.385 | 14.218 | .000 |
| Learn And Can Write Speak AWL | | 1 | 5.230 | 5.579 | .021 |
| Course Focused On AWL | | 1 | 12.053 | 11.419 | .001 |
| AWL was Enhanced within Skills | | 1 | 4.275 | 3.418 | .068 |
| AWL is important for EFL | | 1 | 10.047 | 14.360 | .000 |
| AWL is important for PostGrad | | 1 | 5.176 | 7.141 | .009 |
| I've Been Taught AWL in Class | | 1 | 7.251 | 8.308 | .005 |
| AWL I learned Was Enough | | 1 | .589 | .357 | .552 |
| I Wished I Had More AWL | | 1 | 3.075 | 2.972 | .089 |
| I Had AWL out Class Set By Teacher | | 1 | 1.026 | .714 | .401 |
| I learned AWL out Class By My Self | | 1 | .518 | .402 | .528 |
| Questions: | | 1 | .313 | 1.343 | .250 |
| I Used Text Books to study AWL | | 1 | 1.002 | 5.447 | .022 |
| I Had AWL in Class | | 1 | .004 | .034 | .855 |
| Had Homework By Teacher | | 1 | .239 | .728 | .396 |
| No AWL study This Term | | 1 | .060 | 1.215 | .274 |
| Error | | 164 | | | |

### 6.4.2. Teachers' feedback regarding the AWL learning.

This section explores teachers' views about learners' motivation, their observation about learners' needs concerning the academic words or what they express regarding learning the AWL. It also explores teachers' recommendations regarding what they think achievable during a term from the AWL list, and lastly teachers' personal beliefs that may influence learning of the AWL by their students.

The discussion of the results addresses the two parts of the open-ended questionnaire (beliefs and practices) explained in section 6.2.3.

Seventeen teachers (5 from PY Saudi and 12 from SUni1) participated in the study. As explained in section 6.3., due to the number of respondents, the results are analysed and presented qualitatively.

#### *6.4.2.1. Teachers' Beliefs regarding AWL teaching:*

The questions about beliefs aimed to confirm the notion addressed earlier in section 1.4: that educators favour implicit lexical education over explicit instruction to advanced students due to their perceived learning capabilities. The three statements asked teachers what they thought regarding focusing on the AWL during advanced EFL courses. Determining teachers' feedback and recommendation is thought to have some influence on students' AWL learning, as discussed in section 6.1 above. The results confirmed that some teachers do believe that advanced learners learn vocabulary better incidentally, and that there is no great need for lexical instruction. In fact, half of the respondents did not support direct instruction of the AWL to their advanced learners. The three statements and the quantity of respondents are found beside each response below in Table 6.10:

**Table 6.10 Teachers' beliefs about the AWL teaching to advanced learners**

*1. PY/presessional programmes should have a course that focuses specifically on academic vocabulary.*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | Strongly agree | 6 | Agree | 1 | No opinion | 8 | Disagree | 0 | Strongly Disagree |

*2. Learning academic vocabulary within the course curriculum (language skills and input) is enough.*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Strongly agree | 5 | Agree | 1 | No opinion | 10 | Disagree | 1 | Strongly Disagree |

*3. Instead of allocating some of the schooling time to teach vocabulary, it is better to enhance more important language skills (for example extra writing).*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Strongly agree | 8 | Agree | 0 | No opinion | 8 | Disagree | 0 | Strongly Disagree |

Although statements *one* and *two* should logically have two opposite responses, and, therefore, should have the same count in each direction of the agreement scales, the *disagreement* boxes in the second statement had more ticks than expected. In other words, participants who ticked '*disagree'* are expected to tick '*agree'* in the second statement. Further investigations of the original sheets showed that some guesswork or randomness might have taken place in two of the questionnaires. The participants ticked '*disagree'* for all the three statements.

To conclude, it can be inferred from the some teachers do believe that explicit or direct vocabulary teaching is not necessary for advanced learners. However, approximately half of the teachers recommended the instruction of the academic words to advanced learners. The views of teachers from both sides are explained and elaborated in their comments discussed in the following section.

### *6.4.2.2. Classroom practices and recommendations*

With respect to the importance of vocabulary and specifically the AWL on other language skills, and on reading comprehension discussed in Chapter *2* (e.g. sections 2.2 and 2.3.2.2.), some of the teachers' feedback indicated little emphasis on the AWL in real

classroom practices. Some of the teachers, in fact, completely ignored the AWL in their classes. The majority of teachers either covered some of the list in class or encouraged their learners to work on it independently. Here is the teachers' description:

**Table 6.11 Teachers' description of AWL learning in their classes;**

| *4. How would you describe the learning of Academic vocabulary (specifically the Academic Word List or AWL) that took place in your class?* (more than one option is possible) | **Total 17** |
|---|---|
| A) I covered some of it in class time | 9 |
| B) I covered most or all of it in class time | 1 |
| C) I encouraged students to cover most or all of it in their time | 8 |
| D) I was not aware of such a list for this programme | 1 |
| E) There was a list but I decided not to work on it | 2 |

*Teachers' report on learners' motivation regarding AWL learning:*

Item *5* of the questionnaire aimed to explore learners' motivation about learning the AWL. The open-ended question (see Appendix J and the responses in Tables 6.13-6.15) asked teachers to describe briefly how motivated their learners were, then comment on their motivation. 5 teachers stated that their learners were highly motivated and the majority thought their motivation was fair.

**Table 6.12 Teachers' descriptions of learners' motivation (1);**

| *5. How motivated were your students to learn academic vocabulary* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5 | Highly | 10 | Somewhat | | Not sure | 1 | Little | 1 | Not at all |

Since the sample size of the respondents was relatively small, it was thought to include all of the comments, and then discuss some of the relevant views. The comments were divided into two tables according to commentators' views regarding their beliefs or descriptions of their learners' motivation.

As seen from the above sample statistics in the previous items, some teachers did not fully support the direct teaching of the academic words to advanced learners. The following comments illustrated in Table 6.13 below justify their thoughts.

**Table 6.13 Teachers' descriptions of learners' motivation (2);**

| Participants | *motivated* | Comment on learners' motivation to learn AWL during course |
|---|---|---|
| **J-PY** | not at all | ''learners have high level of English and lengthy vocabulary and their main focus are academic writing and spelling'' |
| **P-PY** | somewhat | ''it became routine for them hence little motivation was needed'' |
| **PY1** | Little | ''these materials and this approach do not produce the best environment for learning and retaining additional vocabularies'' |
| **SU12** | somewhat | General comment<br>''I understand that the AWL covers a limited range of texts, heavily weighted toward humanities and social sciences. The Schmitt & Schmitt book, which focuses on 'general ' academic words - that come from quality newspapers and magazines not specialized journals- does very little in teaching the specific vocabulary that the master students will be dealing with in their study''. |

Teachers' thoughts and reports regarding their learners' lack of interest with the AWL came with different explanations. The first comment by (J-PY) confirms the common notion mentioned earlier in (section 1.4. and 2.5) that some educators in the area of EFL believe that advanced learners have relatively high levels of vocabulary and therefore should focus on other language skills. However, as argued in section 2.2.3., mastering technical or academic words can significantly increase levels of comprehension. Advanced learners with larger vocabulary sizes do not necessarily master those technical terms, as discussed and cited in section 1.3. This was also confirmed using multiple ability measures in the findings of Chapters 4 and 5 in this thesis. Commenter PY1 thought that the (extra) materials used with the treatment groups ''do not produce the best environment for learning and retaining additional vocabularies''. From a one-to-one interview while taking the survey he said he believed that advanced students learn language input better incidentally, reflecting the comment by Folse (2011) mentioned in

section 2.5. Teacher (P-PY) believed that his learners became less motivated as the 'extra' materials became a routine for them. Although the research questions for this Chapter are limited and focused, it was thought that it would be interesting to look at the comments by this teacher with regard to his own learners' achievements. It was found that his learners from one of the treatment group DG, had in fact progressed better than the passive NnDG group, regardless of their lack of interest in the AWL activities as reported by the teacher.

On the other hand, the majority of the teachers reported that their learners were very motivated to learn the AWL, or at least many of them were. The way in which they described their learners' motivation is outlined in Table 6.14 below:

**Table 6.14 Teachers' descriptions of learners' motivation (3);**

| Participants | Comment on learners' motivation to learn AWL during course | |
|---|---|---|
| SU1 | Somewhat | students appreciate the value of the academic words in their further studies and generally keen to learn them |
| SU6 | Somewhat | As with all groups the attitudes varies from person to person with some appreciating its importance fully and others to the extreme not really overly moved to care. I tried to impress on all of them the AWL insignificance. |
| SU7 | Somewhat | dependent on students there was a realization that such vocabulary was necessary |
| SU9 | Somewhat | some of the students weren't as motivated on others, as would be expected |
| SU10 | Somewhat | I thought students realized that it is necessary to widen their range of academic vocabulary |
| SU3 | Somewhat | N/A[1] |
| PY2 | Somewhat | N/A |
| SU2 | Highly | very motivated and realize the importance in their studies and help them to wider their vocabulary in writing |
| SU8 | Highly | they were keen to learn academic vocabulary and most were Master students |
| SU11 | Highly | very often focused on what was an 'academic word' not focused enough on the context and grammar |
| SU4 | Highly | N/A |
| SU5 | Highly | N/A |

[1] N/A indicates that participants provided no comments in this item.

*Teachers' recommendations of feasible AWL targets:*

Item 6 of the questionnaire reports on teachers' recommendation of what they thought was achievable within a term regarding covering the academic list. Many of the teachers seem to agree that about half of the academic word list could be covered during the term with advanced learners. Some teachers in fact said they had covered almost half of the list during one term. Some thought that teaching 15 words was feasible during a lecture/class. One teacher felt covering 25 words a week was feasible. Details of this feedback are reported in Table. 6.15 below:

**Table 6.15 Teachers' recommendations of feasible AWL targets over a term;**

| Participants | Recommendation of feasible AWL targets |
|---|---|
| J-PY | N/A[1] |
| P-PY | The activities over 10 weeks were feasible |
| PY1 | N/A |
| PY2 | I covered the vocabulary in textbooks within the chapters. |
| SU1 | 12/15 per lesson regardless if academic or not |
| SU2 | N/A |
| SU3 | It is much better to learn academic words in textbook. Students can become swamped if they have lists of words. |
| SU4 | Unsure as we taught specific & incidental academic vocabulary. However, I would recommend input of no more than 15 words |
| SU5 | 200-250 |
| SU6 | 250 was covered productive vocabulary increase of 50 on average. 250 is a feasible level unless there's dedicated AWL vocabulary course. |
| SU7 | the presessional course concentrated on approximately 250 words |
| SU8 | 8 reading and approximately 150/200 words |
| SU9 | I'd say about 5 for lesson |
| SU10 | I covered 25 a week so about 250-300 words in 10 weeks term |
| SU11 | 200-300 |
| SU12 | N/A |

[1] N/A indicates that participants provided no comments in this item.

*Teachers' notes about the impact of the knowledge of the AWL on learners:*

The last item of the questionnaire asked teachers to report how their students' learning of the AWL impacted on their skills. The aim of this question was to record any remarks by

the teachers who covered the list with their learners. A large number from the entire population of this study, as shown in Table 6.16 below, noticed that learning the AWL impacted strongly on learners' reading and writing. Some teachers also thought that knowledge of these academic words affected learners' speaking and listening. A detailed investigation that looked at the questionnaires individually, considering all the specific feedback, concluded that teachers who indicated that they covered the list in class noticed a significant impact more often. Some of the teachers who ticked ''Not sure'' stated that they did not teach all 4 skills and, therefore, were not sure about the progress. Few teachers thought that the knowledge of the AWL had little influence on learners' listening and speaking.

**Table 6.16 Teachers' notes about the AWL impact on learners' skills:**

| 7. To what extent you think the words students have learned have influenced their other specific language skills (number of teachers/17)[a] | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Reading** | **7** | o Strongly | **10** | o Somewhat | **1** | o Not sure | | o Little | | o Not at all |
| **Writing** | **7** | o Strongly | **10** | o Somewhat | **1** | o Not sure | | o Little | | o Not at all |
| **Speaking** | **4** | o Strongly | **7** | o Somewhat | **4** | o Not sure | **2** | o Little | | o Not at all |
| **Listening** | **5** | o Strongly | **9** | o Somewhat | **4** | o Not sure | **1** | o Little | | o Not at all |

The Count of feedback is inconsistent with the total number of teachers, which indicates that in some items, some respondents were not carful while filling in those items. This had also occurred in Table 6.10 and explained at the beginning of section 6.4.2.1.

## 6.5. Summary of the chapter

This study explores the general beliefs and also the experiences of learners studying in an advanced EFL programme regarding learning academic words, to identify what could influence their learning of these words. It also, records their descriptions of their learning

needs and in/out classroom practices. A few research gaps concerning *BALLI*[7] were identified, and some discussions were put forward regarding the absence of specific *BALLI* studies that explore vocabulary learning only (see the discussion sections 2.2.4. and 6.1.). Teachers' feedback was also collected to record their opinions about teaching the AWL, and also to gain a general description of their students' learning. The study used a short 17-item questionnaire to survey the learners, and an open-ended questionnaire to survey the teachers. The participants of this study were all learners and some of their teachers in the *First* and *Second* studies discussed in Chapter 4 and Chapter 5. The questionnaires ask the participants to fill in some of their personal information, and then write their feedback beside the given statements. Many issues regarding background information of all learner participants were identified and considered in the analysis of effects that may influence learners' responses.

The results show that the majority of advanced EFL learners report their needs for learning the AWL in their intensive English programmes, and many learners appreciated the emphasis on the AWL in their courses. Learners generally believe the AWL is important for their studies and that the academic words they learned during their programmes were not enough, so 'they wished they had learned more'. Many learners, as discussed, declared their independent learning of the academic list. Further investigations that considered the differences among participants were conducted to identify any significance difference in the feedback according to participants' background. The variables discussed were: groups, UK vs. Saudi learners, level of study, major of study, level or test of proficiency and years of English EFL study. The three main variants that

[7] *Beliefs About Language Learning Inventory*

show significant differences, and were thought to influence learners' views were: groups, UK vs. Saudi learners, and Saudi DG vs. NnDG. Learners of the treatment group in the second study (Saudi DG group) varied significantly from all other learners in responding to statements about the actual descriptions and/or their appreciation concerning their explicit AWL learning. The descriptions of in-class/out-of-class AWL learning, and the appreciation of learning the academic list was significantly higher among the Saudi DG learners. The variation in responses between the two Saudi groups (DG and NnDG) was also significantly in favour of the AWL explicit extra treatment the DG learners were exposed to. *UK vs. Saudi learners* variant clearly overlapped with the variant *graduate/postgraduate* status, since all of the UK students were postgraduates and Saudis were mostly undergraduates. It has a significant effect on learners views; however, the difference between groups was limited to two statement item (items 11 and 12 as discussed), which were subjects for amendment in the original questionnaire, plus a third item (item 6) which was specific for one of the UK groups not all UK participants. The remaining variables showed either insignificant variations or showed similar interaction to the main effects above.

Regarding teachers' beliefs, feedback and description of practices, and what they think is achievable during a single term from the AWL list, the responses showed that some teachers believe that advanced learners are more likely to learn words implicitly and, therefore, there is no great need for direct lexical instruction. Teachers' descriptions of the actual practices state that some of them, in fact, completely ignored the AWL in their classes. However, the majority of teachers either covered some of the list in class or encouraged their learners to work on it independently. 5 teachers out of 17 described the

motivation of their learners to study the AWL as ‘‘high’’, the rest thought the motivation was moderate. Lastly, teachers’ responses varied in describing how many of the academic words they think feasible to teach during a term. Obviously, their individual responses to this question reflect their beliefs stated earlier in other parts of the survey. Generally, their responses state that 12-15 words per lesson is reasonable, and some say between 150 to 300 words can be learnt during a term. One teacher actually felt that covering 200-300 words during the 10 week term was feasible.

It is important to point out that, while many teachers in this study felt that implicit or incidental learning of vocabulary was the best approach used to teach advanced learners, the majority of learners participating in the study expressed their needs for explicit vocabulary teaching. Further implications from this study are discussed further in Chapter 7 and 8.

# 7. Chapter Seven

# General discussion, summary of the major findings and implications, and conclusion

## 7.1. Overview of Chapter Seven.

This chapter presents a summary of the major findings of the three different studies presented in the previous three chapters of this thesis. Each chapter provided a specific discussion of the results of each experiment in detail. The current chapter draws the research together and discuses of the three investigations generally, presenting a summary of the main findings. The discussions include identifying connections between the findings, as well as to the previous vocabulary acquisition research theories and practices reviewed earlier in the thesis. The chapter considers the practical pedagogical implications and ideas drawn from the findings of the studies, and highlights some of the main contributions to the area of SLA vocabulary learning, and assessment.

The discussions are structured in the form of questions and answers, focusing on the main findings.

## 7.2. Summary of results

This thesis aimed to explore the learning of academic words by advanced EFL learners, using multiple tests to assess their knowledge. As discussed, there is a shortcoming in the SLA literature in determining an accurate level of lexical mastery among advanced EFL learners. Many arguments were presented and limitations discussed regarding the targeted words used in some of the studies (as discussed in section 2.3.2.1), types of measures used for the vocabulary assessment, and some concerns were voiced about research design and the ability to generalize some of the results (see the previous discussions in sections 2.3.2.1 and 2.4.2). As well as this, there is some uncertainty about the most efficient approach to vocabulary learning among advanced EFL learners.

In response to the issues raised in the previous chapters, and to serve the purposes of this thesis, three studies were designed to answer three main queries:

1- How does the depth and breadth of academic vocabulary (as measured by the AWL) of learners at an advanced stage of their pre-university studies change during their preparatory programme?

2- Which approach to vocabulary teaching (incidental/explicit) is likely to lead to the most significant gain?

3- What are learners' and teachers' perceptions and beliefs about vocabulary teaching at this level of proficiency?

As explained before, these main questions suggested further queries and hypotheses which discussed; the level of academic and general vocabulary of learners, rate of development over the targeted durations and the correlations between the different abilities of the lexical knowledge.

The general results from the three studies as shown previously indicate that advanced learners on presessional programmes show higher levels of *Receptive* knowledge of the AWL, as they could recognize at least two thirds of the academic list. Some lower achieving learners at this level of English and this stage of the academic/degree of study could recognize less than 30% of the list (see the discussion and implications below). The *Productive* ability of the AWL was low. Learners were able to produce only about 20% of the academic list. Some high achievers failed to produce more than 50% of the AWL. In terms of improvement during a 12-week presessional programme period, the learners increased their AWL receptively by a range of 34.2 words, and productively at 90.7 words. There might be a *ceiling* factor regarding the receptive knowledge, since the

increase in the gains was unexpectedly lower than the productive knowledge as explained before. There were some lower achievers who failed to produce any words from the AWL, and also none of the learners was able to produce more than half of the list at the end of the course, as mentioned. This could suggest that relying on academic exposure only on presessional programmes (i.e. with advanced learners) might not be sufficient to cover the entire academic list. Concerning the development in the general vocabulary size of the learners, this had increased from about 4842 words to 5231 words at the end of the programme. The different tests were found to correlate significantly with each other.

Concerning the second main question of this thesis and its additional queries, it was concluded that advanced learners exposed to explicit extra AWL treatment activities improved their scores more than learners exposed to the academic list implicitly during the intensive English course. In the receptive AWL test (VLT), the variation in terms of improvement between the two groups (DG or treatment group, and the NnDG control group) was not significant but there was a clear trend in favour of the treatment group. Regarding the results of CATSS, the increase in the improvement of gains among all learners was far more obvious than in the VLT, and the DG learners increased their CATSS scores significantly more than the NnDG. The general vocabulary size was not expected to increase, nor did learners of the two groups vary in terms of improvement, simply because the treatment and the academic language exposure might not directly contribute to this knowledge. In fact, as explained before, the scores have unexpectedly decreased. A follow-up study was conducted to investigate whether the adjustment made to the original test caused this decrease, but the results did not confirm this claim. The

variation between the groups concerning the AWL lexical profile in writing was not significant, but some indirect indications were concluded and discussed section 5.4.2.3.

The last research question of the thesis investigated learners' and teachers' feedback and beliefs regarding the AWL learning. The results show that the majority of advanced learners recognised their need to learn the AWL on the intensive programmes. Many learners appreciated the emphasis on the AWL on their courses, and many wished they had more. They believed that the list was important for their studies, and some spent time on self-study of the AWL. Further investigations that considered the background differences between learners were conducted, to reveal those factors that might have affected learners' feedback. The variables explored were: groups, UK vs. Saudi learners, level of study, major of study, level or test of proficiency and years of English EFL study. Generally, learners' responses were similar apart from in some items where learners varied in descriptions of actual classroom practices. Mainly, as expected, learners of the Saudi DG (treatment) group had given different descriptions from the other groups regarding their experience of learning the AWL, and their appreciation regarding the emphasis on the list in classroom and out of classroom. Concerning teachers' beliefs and feedback, the results confirmed that some teachers do believe that advanced learners better learn words implicitly, and, therefore, that there is no great need for direct lexical instruction. Some teachers completely ignored the AWL in class. The majority of the 17 teachers either covered some of the list or encouraged their learners to cover it independently. In terms of learners' motivation, 5 teachers described their learners as ''highly motivated'', the rest thought the motivation level was moderate. Teachers had

varied opinions about feasible AWL targets during the term. Generally, their responses varied between 150-300 during the term (10-12 weeks) or 12-15 per lesson.

## 7.3. Major findings

The results of the current research demonstrated positive findings regarding vocabulary learning, teaching and assessment. The previous three chapters have provided some evidence of the actual level of the academic vocabulary of advanced learners in their final stage of EFL study, using multiple measurements to assess their knowledge. The studies also presented some evidence about the both the receptive and productive level of AWL development on typical presessional courses. The thesis compared implicit vs explicit (or enhanced) vocabulary teaching to advanced EFL learners, and evaluated the different outcomes using multiple measures. Learners and teachers were given an opportunity to express their insights regarding vocabulary learning methods, vocabulary learning needs and achievable objectives, and the beliefs which may influence their lexical learning. In the following sections, the major findings are discussed and linked to the previous literature.

### 7.3.1. Importance of the AWL for advanced learners

The importance of vocabulary for language proficiency has been extensively discussed in this thesis. Results from Chapter 6 confirmed that leaners declared their need to learn academic words. Learners' beliefs indicate that the learning of the AWL is essential for them, and they realize its importance for their graduate studies. This finding confirmed that of Leki and Carson's (1994) in their investigation of learners' feedback about EAP writing instruction needs. In their open-ended survey which focused on writing revealed that learners express their needs to learning vocabulary, as discussed in section 2.2.4. The

current findings also correspond to the limitations of other studies discussed earlier (in section 2.2.4.), which broadly explored learners' perceptions and beliefs about what influences their learning in general second language learning, i.e. this data was specific to explore learners' beliefs about learning vocabulary only. The data also showed, from the teachers' perspectives, that many learners were highly motivated to learn the AWL. This seems to contradict the notion, discussed earlier, that many advanced EFL learners favour implicit vocabulary learning. In fact, the results showed that some teachers ignored the focus on vocabulary with their advanced students, giving the justification that it might be less important compared to other language skills on intensive programmes. The results confirms the comment of Foles (2011) that ELT pedagogy has favoured 'the natural approach' (implicit learning within communicative skills), since native speakers do not learn most of their vocabulary through explicit instruction. However, with respect to learners' perceptions about vocabulary teaching discussed above, the current study found that it is not always the case. Learners realize the importance of vocabulary for their academic studies. Statistically speaking, some results from Chapter 4 showed that knowledge of the AWL (receptive and productive), correlates with the lexical choice in producing free texts (i.e. affects the range of academic words produced in free writing). This issue has been discussed by Laufer and Nation (1995) who claimed that the lexical frequency profile presented similar figures for pieces of writing by the same writer. They used the old University Word List List (Xue and Nation, 1984), which has been updated by the AWL (Coxhead, 2000). Meara (2005) conducted a follow up study and raised some concerns about the generalization of Laufer and Nation's study regarding the correlation between their receptive test and the productive lexical frequency profile in

writing. As discussed before, one of his logical arguments was concerning the uncertainty of the direct relation between passive and active vocabulary measurement. He believed that to reach a confident generalization, very large groups comprising students with a wide range of vocabulary size should be investigated. Laufer (2005b) responded to the criticism stating that LFP does not tell us whether learners can produce certain words when prompted to do so, but rather tells what proportion of frequent vs. infrequent vocabulary they choose to use in their writing. The current thesis found a positive correlation between productive and receptive tests with the lexical frequency profile in free writing among advanced students, supporting the findings by Laufer and Nation (1995), and the claim by Laufer (2005b).

Lastly, concerning the importance of the AWL for advanced learners, the data from Chapters 5 and 6 show that EFL learners at their final and most advanced language learning stage had lower mastery levels regarding the AWL than expected. With such a level of knowledge, learners are likely to struggle in their academic communications. The following section discusses the actual scores and abilities of advanced students regarding the academic word list.

### 7.3.2. Academic/general vocabulary levels of advanced students on EAS/EAP programmes.

The learners' scores as outlined in Chapters 5 and 6 show that many students have low mastery levels of the AWL in this stage of language proficiency. Since the current thesis applied multiple measures, it was found that advanced learners had a weaker ability to productively recall the AWL than expected from learners of such proficiency level. Regarding the vocabulary overall size, the current thesis found that learners' vocabulary

size varies from approximately 4842 to 5231words on UK presessional courses, and from approximately 4627 to 5424 words among Saudi graduates at advanced English foundation level. Nation (2006) recommended the 8000-9000 words size level for L2 learners aiming to study at the academic level. Some researchers thought this figure seemed challenging or, as they described, a 'safe' recommendation (see for example Laufer and Ravenhorst-Kalovski, 2010). Laufer and Ravenhorst-Kalovski suggested two figures as a response, with an optimal 8000 words level for higher lexical coverage of texts, and a minimal level of 4000 to 5000 words for a modest lexical coverage. The findings of the current thesis show that the latter figure by Laufer and Ravenhorst-Kalovski (4000-5000) seemed to reflect more accurately the reality of vocabulary all learners at this level of advanced EFL study. Of course, this is based on scores of XK-lex measure of vocabulary size.

However, Nation recommended his figure (i.e. 8000-9000 words) in order to ensure a 98% level of comprehension of texts. With respect to this, the current research found that even learners with around 5000 words level do not necessarily know the 570 academic words contained in the AWL. No learner produced even half of the academic list (based on CATSS), and the majority knew only one-third of the list. Learners therefore may struggle with academic texts even with this level of vocabulary size. As seen previously, the list together with the 2000 GSL words cover about 90% (GSL= 80%, AWL=8.5%-10%) of the running words in academic texts. This additional percentage is very significant for text comprehension. As discussed, Webb (2010) found that the pre-learning of the less frequent 10 words in a TV genre represented text coverage of about

0.70% to 3.91%. This therefore means that the knowledge of the AWL words that frequently appear in academic texts affect comprehension significantly.

Nonetheless, the 'best' or 'achievable' vocabulary size recommended for advanced EFL learners demands more investigation. There are two figures recommended for advanced EFL learners in order to reach higher levels of comprehension of texts. The higher figure (8000-9000 words) may seem very challenging, and the minimal level (4000-5000) may seem achievable but represent a lower comprehension level (i.e. 95% coverage as found by Laufer and Ravenhorst-Kalovski 2010). The current thesis speculates about what the knowledge of the AWL could add to the comprehension levels to students with around 5000 words. In other words, since;

a) the 8000 words figure is expected to represent a high coverage of about 98%,

b) the 5000 words represent a modest lexical coverage of about 95%, and

c) the current research found that advanced learners have a weak knowledge of the AWL even with their 5000 words size;

The question must be how much could the percentage of lexical coverage increase if learners knew the AWL plus their 5000 words level. The results of Chapter 4 gave some indication of the effects of the AWL on learners and showed that the scores in both of the two different tests impacted positively on the lexical choice in free writing. Again, this reinforces the great need for teaching the AWL to advanced learners.

### 7.3.3. Explicit or implicit academic vocabulary learning

Schmitt (2008b) commented on vocabulary teaching that there is no 'right' or 'best' way to teach vocabulary as each 'specific learning situation' depends on many factors e.g.

level of students and targeted words, and time allocated for learning methods. In response to one of these 'specific learning situations' i.e. advanced learners on intensive English programmes, the current thesis concludes that explicit or direct vocabulary instruction resulted in better results than implicit teaching. In fact, the direct instruction approach is favoured by many learners on intensive courses. Many teachers described their learners as highly motivated or at least show some motivation to learn academic vocabulary. By contrary, some of the data presented in the current thesis found that some teachers have views against the explicit teaching of vocabulary, and believed that advanced learners are capable of acquiring words from academic exposure, and that there is no great need for lexical instruction. In practice, the results of Chapter 4 reveal that relying only on academic exposure during the contact hours of the presessional course, and right before their academic studies, was insufficient to cover even half of the AWL. Also, since the current study used multiple testing, it found that even high achievers with a level size around 6000 words could not even produce one-third of the AWL at the end of the presessional course. This reinforces the importance of explicit instruction of the AWL, and that the larger vocabulary size a student has does not necessarily mean mastery of the 570 academic words. The findings of Chapter 5 showed that the group of learners who were exposed to the extra activities that enhanced the AWL scored higher than the group who received normal EFL instruction during the intensive English programme. Of course, it should be noted that this extra exposure was not enough, since even the best performing students still knew less than half of the AWL at the end of their programme.

Lastly, AlSaif (2011) in his analysis of twenty course books introduced to Saudi schools, found that only 3800 word families are presented, from which 2800 word families belong

to the most frequent 5000 words. His analysis of teacher-talk during classes revealed that it was mostly below the 2000 words level. Horst (2005) argues that in order to learn the 231 (out of 570) most common academic words, learners must read 28 graded readers of level 5. AlHomoud (2007) argues that direct teaching should enable students to *notice* new vocabulary, either contextualized or decontextualized; whereas extensive reading and listening should help them encounter this vocabulary frequently. This, when we take into the account the superior performance by the explicit group (DG group discussed in Chapter 5) and the results regarding learners' attitudes toward vocabulary teaching (discussed in Chapter 6) strongly suggests the importance of explicit teaching of vocabulary to advanced learners and supports the findings of this thesis.

### 7.3.4. AWL feasible targets over a university term.

As mentioned earlier, the academic vocabulary list consists of 570 word families that appear frequently in academic texts, and represents about 10% of each academic English text. Some of the results of this research showed that advanced learners on presessional courses recognized two thirds of the list, but were able to produce less than 30%. Some learners, who were considered to be advanced, failed to produce any of the AWL. Learners were able to recognize (VLT scores/receptive kowledge) 433 words of the AWL at the beginning of the course, and increased their vocabulary to 467 words, calculated on the basis of a formula set by Schmitt and AlHomoud (AlHomoud, 2007). The increase was equal to 34.2 words a term, and less than a word a day (.75). A ceiling effect might have affected the degree of improvement of this measure since learners' scores reached the maximum score. In terms of learners' ability to produce the AWL (CATSS scores),

they were able only to produce 98 words, and gained about 90.7 words during the course. This represents a rate of 2.01 words a day.

The results of Chapter 5 involved two learning groups as discussed earlier. Learners from both groups were able to recognize at least 403 words. The improvement in the receptive knowledge of the AWL during the term for both groups was not high, but the active (treatment) group improved significantly more than the control group, increasing knowledge to 422 words. Concerning productive knowledge, all learners shared almost the same figure of about 114.2 academic words. Learners of the active group gained 96.4 words, and the passive (or control) group of learners gained 38 words.

Chapter 6 investigated what teachers believe is feasible to cover from the list during the term. They indicated that a gain of 12/15 words a lesson is feasible, and some think that an increase of 150 to 300 words is achievable within a term. Some teachers felt covering 25 words a week (based on two lessons) during the ten weeks of the programme was reasonable. Milton (2009) and Milton and Meara (1998) report from many studies conducted in different contexts and different countries that EFL learners gain between 462- 600 words annually as the highest estimates, and in the lowest estimates between 200 and 400 words. They report from different studies that learners gain about 4.4 words per classroom hour. Many figures were also reported earlier about vocabulary gains from extensive reading (AlHomoud, 2007; Horst, 2005; Pigada & Schmitt, 2006). These figures, of course, are the rates of general word gains. However, as reviewed earlier in the chapter, there is an absence of research focusing on general AWL gains among advanced learners. The current thesis showed the gains regarding the academic words (receptive

and productive) among advanced learners during a term, in two different research contexts.

### 7.3.5. Need for multiple testing.

A great body of research has discussed the aspects of vocabulary knowledge and what knowing a word entails (e.g. Nation, 1990, 2001; Richards, 1976; Schmitt, 2010a; Webb, 2008) as discussed in section 2.3.1. Depending on a single measure is unlikely to provide a clear indication about learners' knowledge of the word. Hence, scholars always emphasize the importance of applying multiple measures in studies and experiments investigating vocabulary growth due to the complexity of word knowledge (Laufer, Elder, Hill, & Congdon, 2004; Nation, 2007b; Read, 2004; Schmitt, 2010a). Research in vocabulary testing has discussed the correlation between the general receptive size of learners and their productive size. The receptive size is always larger, and the difference between the two dimensions increases with non-common words (Webb, 2008). Milton (2009) concludes that many studies show that productive knowledge is 50-80% of the receptive knowledge. Schmitt (2010: 152) comments ''it might be reasonable to assume that a demonstration of productive knowledge also implies receptive mastery, based on research which shows that this is generally true'' however, "the real danger is making generalizations in the other direction''. i.e. over-estimating learners' productive knowledge based on their performance on a test of receptive knowledge. The current study, however, focused on the academic words and showed the relationship between learners' recognition abilities of the AWL and their production levels. As reported in the previous section, learners showed that they could recognize an equal of at least 407 words out of the 570, but they were only able to actually produce 90.2 words. With respect to

Milton's conclusion about general words size, this could mean that the percentage of the productive knowledge of the AWL among advanced students is 22.3% to 39.5 % of the receptive knowledge. Of course, the percentage is based on the academic vocabulary which is the main focus of this study.

Lastly, as identified, learners showed higher receptive knowledge reaching the maximum scores, but showed very weak productive knowledge. If the current study had depended only on a single measure, it would not be able to capture the accurate vocabulary levels among learners, and may have given overestimations about their mastery of words. This reinforces the importance of applying multiple measures to assess learners' AWL abilities, and shows how it is unwise to overestimate productive abilities relying on receptive abilities of the AWL, their overall vocabulary size, or their proficiency. Assuming that learners are able to produce academic words based on their advanced levels of proficiency or large vocabulary size is unwise. Also, detecting learners' improvement over a university term and drawing some empirical implications and conclusions would not have been possible if a single measure was used in the assessment (e.g. the difference of the improvement in the receptive test was not as obvious as in the productive test).

## 7.4. Research and pedagogical implications

The importance of this research lies in its significance in invalidating some of the uncertain logic, comments or common perceptions (see the discussion 2.3.2.2) regarding some of the current applications of teaching of vocabulary to learners in advanced EFL programmes. The three investigations theorise to the following important research and pedagogical implications.

1- EFL learners with an advanced level of proficiency and/or larger vocabulary size do not necessarily mastery the entire academic words list. Knowledge of the list significantly affects learners' academic communication, as concluded by Webb (2010), and suggested by the findings of the current investigations. Many advanced learners were found to be struggling with the production of the AWL. Therefore, assessing their learning needs and emphasizing the AWL is recommended.

2- Using multiple assessment of vocabulary is extremely important. Learners in the current studies seemed to recognize most of the academic words. However, their productive abilities of the AWL were very weak. Applying multiple measures is hugely important for practitioners and researchers investigating lexical development in order to diagnose all mastery levels.

3- Assessing advanced students' general vocabulary size and the academic words is by itself very useful in making learners aware of their lexical abilities at this stage of language proficiency. It is also a useful motivational tool. Assessing learners' AWL mastery levels at the beginning of their programme would give them a picture regarding their actual abilities, and therefore encourage them to focus their learning and work on their own weaknesses.

4- Relying only on academic exposure in advanced EFL courses that allow implicit learning of academic vocabulary might not be sufficient for satisfactory AWL acquisition. Horst (2005) mentions that in order to learn the 231 (out of 570) most common academic words, learners must read 28 graded readers from level 5. Unfortunately, this amount of reading might be very challenging and

perhaps impossible to maintain on presessional or intensive EFL courses. Therefore, enhancing the AWL via direct tasks and exercises would be more efficient, and probably result in reaching the desired objectives more quickly. Of course, it must be noted that although the current research found that direct enhancement of academic vocabulary gave better AWL improvement than the implicit approach, the direct approach was still insufficient to master the entire list.

5- Explicit/direct teaching of vocabulary resulted in greater and faster gains as found in Chapter 5. Learners who had extra activities that enhanced the AWL improved 60% more than learners who were exposed to normal academic teaching only. Thus direct teaching of the AWL is highly recommended.

6- It could be concluded from the three studies that teaching at least 250 family words (out of 570 words) is achievable during a university term. One teacher claimed to cover 25 words a week over 10 weeks. Many teachers believed that 12/15 words is achievable in a lesson by advanced learners. As mentioned earlier, Horst recommended reading 28 books of level 5 of graded readers to learn the most common 231 academic words. Explicit teaching of the AWL could lead to achieving the same learning targets, but in less time. Of course, more effort and teaching should be made to in order to cover the entire academic list.

7- Exploring learners' beliefs and feedback that may influence their vocabulary learning is almost non-existent in SLA research. Most research into vocabulary has been included within broader investigations of language learning beliefs.

The results of Chapter 6 confirm that explicit vocabulary teaching is appreciated by advanced learners. This invalidates any incorrect assumptions that may lead teachers to believe that their advanced learners demand less explicit teaching of vocabulary due to their high levels of proficiency, as discussed in section 2.3.2.2. Some of the findings of Chapter 6 confirm the occurrence of this view and that some teachers assume that explicit tasks are less preferable by advanced learners. As reported, some teachers think that advanced learners find direct approaches boring compared to communicative tasks, and therefore, learners become less motivated in time. However, leaners' perceptions confirmed that this is not always the case.

8- Learners sometimes appear to be less motivated to learn vocabulary via explicit tasks, as noted by a teacher of one of the DG (treatment) groups in Chapter 6. However, further investigations showed that learners of that specific group increased their vocabulary range more than learners of the NnDG group, regardless of any signs of poor motivation. Therefore, teachers should try different ways of motivating their learners and carry on with efficient approaches, even if learners seemed less interested in certain activities.

9- Data in Chapter 5 showed that some Saudi learners joining intensive EFL courses have poorer vocabulary size than expected from learners of such a level of proficiency. Some learners had a vocabulary size of less than 2000 words. This surely means that they will struggle in their future studies, and may not even pass the PY programmes. Teachers should set up some lexical teaching plans for such weak students. They should design some extra reading (e.g.

graded readers) assignments for them, plus some direct vocabulary exercises that target the most frequent 2000 words.

10- Due to the complexity of vocabulary mastery, discussed extensively in this thesis, multiple measures were recommended for better judgment of and decisions about vocabulary levels and learning plans. However, as discussed in Alothman (2013), applying multiple measures to investigate vocabulary development is likely to interrupt much of class time, which might affect data collection. Schools and/or teachers might not agree to their learners taking many tests in one day or lecture. If the tests are distributed over many lectures, data might be lost or become invalid at the end of the study, especially with long-term investigations. Therefore, researchers are advised to be aware of these issues, and are recommended to be cautious and well prepared when planning their experiments and investigations of vocabulary improvement, considering those issues. For example, researchers could create online versions of the tests and look to deliver them out of class time – ideally in the learners' own time. To achieve this, the value of assessment must be stressed to avoid cheating and practical and personalized feedback provided.

### *7.4.1. Further observations and implications*

1- Scholars always recommend a combination of the two learning methods of vocabulary teaching; explicit or direct teaching and implicit or incidental learning. A good vocabulary plan should include both approaches to teaching to ensure enhancing and processing as much aspects of vocabulary knowledge as possible. The direct enhancement/teaching of vocabulary explored in the current

thesis could support incidental AWL learning within academic contents as it allows earlier noticing of words, as discussed in sections 1.4 and 2.3.2.1.

2- There are many textbooks now available to teach the AWL; such as (Campbell, 2012; Huntley, 2006; McCarthy & O'Dell, 2008; Richmond & Zimmerman, 2007; D. Schmitt & Schmitt, 2005). They provide materials, exercises and contexts that enhance much knowledge aspects of the academic words. Educators are encouraged to include such textbooks in their curriculum or assign learners to study from them independently.

3- Teachers and learners can also make use of the materials available on the Internet that enhance the academic words list. Most learners also carry smart phones and devices in their pockets, and can download many useful software and applications. Teachers could create some mobile apps and/or simply encourage their learners to use the best applications that provide exercises for learners. The application '*The Academic Words'* found in https://itunes.apple.com/us/app/academic-word/id601704455?ls=1&mt=8, and developed by the author of this thesis is among those apps that target the AWL at Arabic speakers on the PY programmes. The advantage of such applications is that learners can access such interactive learning tools on their smart phones anytime anywhere.

## 7.5. Limitations

The current investigations must consider some limitations of the results which were mostly beyond the control of the research procedures.

1- Although the current thesis used multiple tests to assess vocabulary, it did not include other aspects of word knowledge. The tests that targeted the academic words consisted of a recognition test and a controlled production test. The two tests gave two different results; however, testing the ability to produce words comes in different forms. Further investigations should also include more aspects of word mastery (as discussed further in section 7.6). Furthermore, although the current thesis included a relatively large sample size of participants, the correlations that were found between the AWL tests and the vocabulary size test (XKlex) should be taken with caution. XKlex is criticized by some researchers as giving overestimations of level size, as discussed in 2.4.4. Despite this criticism, the test has been validated in major research works and produced consistent results (Masrai, 2009). It was used in the current thesis as it serves logistically.

Another aspect of this limitation, as discussed before in Chapter 5, XKlex presented some unexpected results. Only participants of the second study discussed in Chapter 5 decreased their scores. Although a follow-up validation was made and could not confirm the null hypothesis (i.e. learners were aware of a pattern in the test items), future research should investigate XKlex more thoroughly, especially among Saudi learners.

2- Many students participated in the multiple measures of the thesis; however, very few learners completed the writing tasks, which may give rise to some uncertainty regarding the generalizability of any results. The correlations between the ability to produce academic words freely in writing and the AWL

tests demand further investigations. More samples of writing are recommended to confirm any correlations (see further discussions on the following section). As noted before also, the frequency of using the AWL in writing is uncontrolled regardless of the care taken in the thesis. Further investigations would confirm the indications that made in the current study. As discussed before, further qualitative data was reported regarding the limited sample in writing.

3- Some learners in the *first* study and the second study might have studied the academic words independently during the programmes, and this, would have an effect on their test scores. However, the surveys (which included participants' names) about beliefs and practices (discussed in Chapter 6) reduce concerns about this issue.

4- Concerning the reliability of some of the measures, the accepted minimum level of reliability (0.70) was not reached in three occasions as discussed and noted before. However, the length of some of the instruments, combined with the relatively small and truncated population and the relative homogenous nature of the groups make this inevitable. Therefore, the estimates achieved can be seen as at least acceptable (Punch, 2005).

## 7.6. Suggestions for future research

The current thesis can act as a starting point to motivate future research to follow up on some of its conclusions. Below are some suggestions.

1- The current research used three levels of assessing academic words knowledge; recognition, controlled production and the frequency of free uncontrolled production of

the words in writing. Due to the complexity of vocabulary knowledge, further research is recommended to include more aspects of word mastery (e.g. spoken vocabulary knowledge, knowledge of different forms or derivatives of the words, different measures of productive or receptive knowledge). The 570 words list is an achievable target for researchers to sample some words and follow up with further deeper assessments.

2- The relationship between general vocabulary size and knowledge of the AWL demands further research. Nation recommends not less than 8000 words to reach 98% level of comprehension of academic English reading texts. Laufer and Ravenhorst-Kalovski (2010) responded by suggesting two figures; an optimal 8000 words level for higher lexical coverage of texts, and a minimal level of 4000 to 5000 words for a modest lexical coverage. The current thesis found that the latter figure (i.e. 5000 word level) was closer to the vocabulary range among learners on the presessional programmes participating in this research. However, these learners demonstrated a lack of ability to produce the AWL, while the majority could recognize only two thirds of the list.

3- As discussed, the list covers about 8.5%-10% of the running words in academic texts (Coxhead & Nation, 2001; Nation, 2006). Nation (2006) recommended 8000-9000 vocabulary thresholds to reach 98% of comprehensions of texts, and Laufer and Rozovski-Roitblat (2011) believe Nation's figure was challenging and suggested 4000-5000 words level. Knowledge of the AWL in general was found to correlate significantly with overall vocabulary size. With respect to all the above, the current thesis speculates about how far the comprehension level could increase if learners of 5000 words level learned the 570 academic word list. Further research is recommended to identify the link between vocabulary size and knowledge of the AWL with respect to the comprehension

levels of academic reading. In other words, since the main two studies of this thesis investigated the relationship between the vocabulary size (including the AWL) and reading comprehension, it would be interesting to know how far the performance of the AWL (apart from the vocabulary size) affects comprehension.

3- XK-lex (Meara & Milton, 2006) test of vocabulary size was chosen in this research as it serves better logistically (i.e. easy to administer and requires less of participants' time in this multifaceted research). The test was trialed and validated in major research studies (e.g. Masrai, 2009). However, it was criticized for giving an overestimation of vocabulary size, as discussed above. Future research is recommended to investigate the use of other vocabulary estimates tools such as; Nation's *vocabulary level test VLT* (Schmitt, Schmitt, & Clapham 2001), Laufer's *Computer Adaptive Test of Size and Strength CATSS* (Laufer & Goldstein, 2004) and others as described in Section 2.4.4. Each measure works differently in working out the estimates of vocabulary size of testees. Furthermore, the resluts of XK-lex in Chapter 5 showed unexpected results, since the scores in the posttest decreased. The original test has a pattern in the construction of its items as described before. Although a follow-up study was conducted and confirmed that learners were not aware of any pattern in the construction of the test items, future qualitative and quantitative research is recommended in addressing the issue more thoroughly.

4- The current study had some limitations about the relationship between the receptive/productive AWL abilities and the lexical frequency of the academic words in free writing tasks. Further research is recommended to include longer essays and more writers.

5- Teachers' beliefs and perceptions about advanced EFL learners concerning vocabulary teaching demand some further investigation. Teachers have different views concerning the best approach to teaching vocabulary to advanced English learners. Some teachers seem to agree with the common notion in second language research that implicit teaching within communicative language tasks is favored by learners due to their linguistic ability and being closer to the L1 method of language acquisition. However, the current thesis showed that learners do appreciate vocabulary instruction, and confirm their lexical learning needs. Future research is recommended to apply quantitative and qualitative research studies to explore teachers' beliefs and perceptions further.

## 7.7. Conclusion.

The investigations in this thesis concluded that learners at 'what is considered' the last stage of EFL study were found to have poor ability in producing the AWL, regardless of their high scores in the receptive test. In fact, they were able to produce very little of the AWL, which should question previous assumptions that learners could produce 50-80% of the words they know receptively. Academic exposure during presessional courses resulted in unsatisfactory improvement of the AWL, which suggests that planning more deliberate teaching of the AWL is necessary. Direct or explicit teaching resulted in better gains of the AWL during a university term of an intensive EFL programme. Learners exposed to the extra treatment of the AWL exercises improved 60% more than learners exposed to normal academic teaching only. This clearly reinforces the importance and usefulness of explicit vocabulary teaching to advanced learner, who are widely perceived to be capable of learning language incidentally. Of course, as noted, this improvement is still not sufficient to master the entire academic list. Lastly, the thesis explored a

neglected area of research concerning learners' beliefs which may influence their learning. Investigating learners' beliefs about vocabulary learning has been almost non-existent in SLA research. The current research gained many valuable insights from learners about vocabulary learning. They seemed to appreciate and favour explicit vocabulary teaching over implicit learning. This surprisingly contradicts many assumptions that advanced learners prefer implicit learning tasks due to their high level of proficiency and linguistic capabilities. In fact advanced learners should be taught academic vocabulary deliberately because of their linguistic capabilities, not the opposite. Direct lexical instruction allows more immediate encounter of vocabulary and therefore better and faster progression is expected.

Scholars have debated at length about the issue of vocabulary acquisition, as discussed above. Krashen (1989) argues that 'we acquire vocabulary and spelling by reading'. In supporting his 'Input Hypothesis', he suggested that similar to L1 speakers, L2 learners acquire words, particularly from reading, by focusing on the message not on the words. However, researchers have provided evidence that challenges the claim that reading is the major source of vocabulary acquisition, and that in fact focusing on form gives better results in terms of words gain (e.g. Laufer, 2003; Qian, 1996). Researchers have presented a number of practical limitations that criticize extensive reading programmes as a slow approach to vocabulary size improvement, and a challenging method for beginners. Nation and Wang (1999) proposed graded readers for beginners in response to opponents of adopting extensive reading programmes to extend vocabulary, and offered a suggestion for their practical limitation, i.e. words appearing in graded readers are suitable for beginners. However, as argued before, Horst (2005) states that the AWL was

not frequent in the 37 graded readers (level 5) she scanned. Likewise, Milton (2009) argues that the materials in textbooks do not always address the issue of vocabulary frequency. With respect to all these issues, along with the findings that this thesis provides, it could be comfortably concluded that explicit instruction of vocabulary is more efficient for learners than implicit teaching.

## References


Adolphs, S., & Schmitt, N. (2003). Lexical Coverage of Spoken Discourse. *Applied Linguistics*, *24*(4), 425–438.

Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language*. John Benjamins Press.

Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, *38*(3), 391–401.

Al-Akloby, S. (2001). *Teaching and learning English vocabulary in Saudi Arabian public schools: An exploratory study of some possible reasons behind students" failure to learn English vocabulary*. University of Essex.

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. Urquhart (Eds.), *Reading in a Foreign Language* (pp. 1–27). Longman.

Alemi, M., & Tayebi, A. (2011). The Influence of Incidental and Intentional Vocabulary Acquisition and Vocabulary Strategy Use on Learning L2 Vocabularies. *Journal of Language Teaching and Research*, *2*(1), 81–98.

Alfraidan, A. (2010). *test taking staratigies of EFL Saudi University level*. *thesis*. University of Essex.

Al-Hazemi, H. (1993). *Low level EFL vocabualry tests for Arabic speakers. Unpublished Phd Thesis.* University of Wales, Swansea.

Al-Homoud, F. (2007). *Vocabulary acquisition via extensive input.* University of Nottingham.

Al-Nujaidi, A. (2003). *The relationship between vocabulary size, reading strategies, and reading comprehension of EFL learners in Saudi Arabia*. Oklahoma State University, USA.

Al-Nujaidi, A. (2004). The Impact of extensive reading on EFL vocabulary development. In *TESOL Arabia*. Dubai, UAE.: TESOL Arabia.

Alothman, K. (2009). *Focusing on Academic Vocabulary in an EAS programme*. University of Southampton.

Alothman, K. (2011). Teaching Vocabulary at Preparatory and/or foundation years: what is fast and more effective. In *The 2nd SIELT Conference: Minimum Time, Maximum Output April 8th-9th*. Anadolu University Conference.

Alothman, K. (2013). Multiple testing of vocabulary knowledge: benefits and challenges that may affect practicality. In *Langue Testuing in Europe: May 28, 2013*. Antwerp, Belgium.

AlQahtani, M. (2005). *The use of vocabulary learning strategies by EFL learners at three different educational levels*. University of Essex.

Al-Rajhi, A. (2004). *Joining the Online Literacy Club: Internet Reading among Saudi EFL Learners*. Indiana University of Pennsylvania.

AlSaif, A. (2011). *Investigating Vocabulary Input and Explaining Vocabulary Uptake among EFL Learners in Saudi Arabia*. *Philosophy*. Swansea University.

Astika, G. . (1993). Analytical assessments of foreign students' writing. *RELC Journal*, *24*, 61–72.

Beasley, T., & Zumbo, B. (2009). Aligned Rank Tests for Interactions in Split-Plot Designs: Distributional Assumptions and Stochastic Heterogeneity. *Soph.uab.edu*, *8*(1), 16–50. Retrieved from http://www.soph.uab.edu/Statgenetics/People/MBeasley/Beasley-Zumbo-AlignedRanks-JMASM-2009.pdf

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford Press.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*(1), 101–118. Retrieved from http://ltj.sagepub.com/cgi/doi/10.1177/0265532209340194

Bernat, E. (2006). Assessing EAP learners' beliefs about language learning in the Australian context. *Asian EFL Journal*, *8*(2).

Borer, L. (2007). Depth of processing in private and social speech: Its role in the retention of word knowledge by adult EAP learners. *Canadian Modern Language Review/La Revue*, *64*(2), 273–299. Retrieved from http://utpjournals.metapress.com/index/93w061626p600l11.pdf

Brown, D. (2011). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research*, *15*(1), 83–97.

Campbell, C. (2012). *English for Academic Study: Vocabulary*. Garnet Education.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd ed.). London: Routledge.

Chen, C., & Truscott, J. (2010). The Effects of Repetition and L1 Lexicalization on Incidental Vocabulary Acquisition. *Applied Linguistics*, *31*(5), 693–713.

Ching-Shyang chang, A. (2007). The impact of vocabulary preparation on L2 listening comprehension, confidence and strategy use. *System*, *35*(4), 534–550.

Choudhury, anidya syam. (2010). Teaching Vocabulary in the ESL / EFL Classroom : Central Pedagogical Issues. *Mjal*.

Chui, A. S. Y. (2006). A study of the English vocabulary knowledge of university students in Hong Kong. *Asian Journal of English Language Teaching*, *16*, 1–23.

Chung, T., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, *15*(2). Retrieved from http://nflrc.hawaii.edu/RFL/October2003/chung/chung.html

Coady, J. (1993). Research on ESL/EFL vocabulary acquisition: Putting it in context. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 3–23). Norwood, NJ: Ablex Pub. Corp.

Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition*. *The Cambridge Applied Linguistics* (Vol. 9). Cambridge University Press.

Cobb, T. (2008). Web Vocabprofiler (Version 2.6) [Computer software]. http://www.lextutor.ca/vp/. Retrieved from http://www.lextutor.ca/vp/

Coxhead, A. (2000). A New Academic Word List. *Tesol Quarterly*, *34*(2), 213–238.

Coxhead, A., & Nation, P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252–267). Cambridge University Press.

Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, *42*(8), 895–901.

Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. (H. Daller, J. Milton, & J. Treffers-Daller, Eds.)*The Cambridge Applied Linguistics Series* (Vol. 11, p. 289). Cambridge University Press.

Daller, H., & Phelan, D. (2007). What is in a teacher's mind? Teacher ratings of EFL essays and different aspects of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 234–244). Cambridge: Cambridge University Press.

Das Neves Seesink, M. T. (2007). *Using blended instruction to teach academic vocabulary collocations: A case study*. ProQuest.

Elgort, I. (2011). Deliberate Learning and Vocabulary Acquisition in a Second Language. *Language Learning*, *61*(2), 367–413.

Eyckmans, J. (2004). Measuring receptive vocabulary size. *Landelijke Onderzoekschool Taalwetenschap', LOT, the Netherlands National Graduate School of Linguistics. Utrecht, the Netherlands*.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications Ltd.

Folse, K. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. University of Michigan Press Ann Arbor, MI.

Folse, K. (2006). The Effect of Type of Written Exercise on L2 Vocabulary Retention. *Tesol Quarterly*, *40*(2), 273–293.

Folse, K. (2010). Is explicit vocabulary focus the reading teacher's job? *Reading in a Foreign Language*.

Folse, K. (2011). Applying L2 Lexical Research Findings in ESL Teaching. *TESOL Quarterly*, *45*(2), 362–362.

Gardner, D. (2007). Validating the Construct of Word in Applied Corpus-based Vocabulary Research: A Critical Survey. *Applied Linguistics*, *28*(2), 241–265. doi:10.1093/applin/amm010

Gass, S., & Selinker, L. (2001). *Second language acquisition: An introductory course*. Mahwah NJ: Lawrence Erlbaum Associates.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, *11*(4), 341–363.

Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics*, *24*, 44–64.

Hincks, R. (2003). Pronouncing the Academic Word List: Features of L2 student oral presentations. *Proceedings of the 15th International Congress of*, *28*(1), 1545–1548. Retrieved from http://www.speech.kth.se/~hincks/papers/ICPh03.pdf

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*, 689–689.

Horst, M. (2005). Learning L2 Vocabulary through Extensive Reading: A Measurement Study. *The Canadian Modern Language Review La Revue Canadienne Des Langues Vivantes*, *61*(3), 355–382. doi:10.1353/cml.2005.0018

Horst, M. (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language*, *22*(1), 161–180. Retrieved from http://nflrc.hawaii.edu/RFL/April2010/articles/horst.pdf

Horwitz, E. (1988). The beliefs about language learning of beginning university foreign language students. *The Modern Language Journal*, *72*(3), 283–294.

Hu, H.-C., & Nation, P. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.

Huckin, T., & Coady, J. (1999). Incidental Vocabulary Acquisition in a Second Language. *Studies in Second Language Acquisition*, *21*(02), 181–193.

Hulstijn, J., & Laufer, B. (2001). Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning*, *51*(3), 539–558.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, *17*(1), 23–59. Retrieved from http://nflrc.hawaii.edu/rfl/april2005/hunt/hunt.html

Huntley, H. (2006). *Essential academic vocabulary: Mastering the complete academic word list* (Vol. 10). Boston: Houghton MifflinCompany.

Ishii, T., & Schmitt, N. (2009). Developing an Integrated Diagnostic Test of Vocabulary Size and Depth. *RELC Journal*, *40*(1), 5–22.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition. *Applied Linguistics*, *19*(3), 357–377. doi:10.1093/applin/19.3.357

Joe, A. (2010). The quality and frequency of encounters with vocabulary in an English for Academic Purposes programme. *Reading in a Foreign Language*, *22*(1), 1.

Krashen, S. (1985). *The Input Hypothesis: Issues and implications*. New York: Longman.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, *73*(4), 440–464.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Laurén & M. Nordman (Eds.), *Special Language From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and Applied Linguistics* (pp. 126–132). Macmillan.

Laufer, B. (1997). The lexical plight in second language reading. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 20–34). Cambridge: Cambridge University Press.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, *19*(2), 255–271.

Laufer, B. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. doi:10.1093/applin/22.1.1

Laufer, B. (2003). Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *Canadian Modern Language Review La Revue Canadienne Des Langues Vivantes*, *59*(4), 567–587. doi:10.3138/cmlr.59.4.567

Laufer, B. (2005a). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook*, *5*(1), 223–250.

Laufer, B. (2005b). Lexical Frequency Profiles: From Monte Carlo to the Real World: A Response to Meara (2005). *Applied Linguistics*, *26*(4), 582–588. doi:10.1093/applin/ami029

Laufer, B. (2006). Comparing Focus on Form and Focus on FormS in Second-Language Vocabulary Learning. *The Canadian Modern Language Review La Revue Canadienne Des Langues Vivantes*, *63*(1), 149–166. doi:10.1353/cml.2006.0047

Laufer, B. (2009). Research Timeline: Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching*, *42*(03), 341.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202–226.

Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning*, *54*(3), 399–436.

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, *16*(3), 307–322.

Laufer, B., & Nation, P. (1999). Laufer Productive levels Test. *Language Testing*.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited : Lexical text coverage , learners ' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurance and their combination. *Language Teaching Research*.

Lawson, M., & Hogben, D. (1996). The Vocabulary Learning Strategies of Foreign Language Students. *Language Learning*, (March), 101–135. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-1770.1996.tb00642.x/abstract

Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System*, *31*(4), 537–561.

Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *Tesol Quarterly*, 81–101.

Lewis, M. (1997). *Implementing the Lexical Approach. Putting Theory into Practice.* Hove: Language Teaching Publications.

Llach, M., & Gallego, M. (2009). Examining the relationship between receptive vocabulary size and written skills of primary school learners. *ATLANTIS Journal of the Spanish Association of Anglo-American Studies*, *31*(June), 129–147. Retrieved from http://www.atlantisjournal.org/ARCHIVE/31.1/2009Agustin_Terrazas.pdf

Long, M. (1991). Focus on form: a design feature in language teaching methodology. In K. De Bot, D. Coste, R. Ginsberg, & C. Kramsch (Eds.), *Foreign Language Research in Crosscultural Perspectives* (pp. 39–52). Amsterdam: John Benjamins.

Martinez, R. (2010). Evidence of lack of processing of multiword lexical items in reading tests. In *Paper accepted for presentation at the Language Testing Research Colloquium (LTRC)* (p. April 14). Cambridge, UK.

Masrai, A. (2009). *Measuring the English vocabulary size of Saudi university students: Validating a new 10,000 word vocabulary size test*. Swansea University.

Matsuoka, W., & Hirsch, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities. *Reading in a Foreign Language*, *22*(1), 56–70.

McCarthy, M., & O'Dell, F. (2008). *Academic Vocabulary in Use*. Cambridge University Press.

Meara, P. (1980). Vocabulary Acquisition: A Neglected Aspect of Language Learning. *Language Teaching*, *13*(3-4), 221.

Meara, P. (1990). Some notes on the Eurocentres vocabulary tests. *AFinLA Yearbook*, 1–8. Retrieved from http://www.lognostics.co.uk/vlibrary/meara1990b.pdf

Meara, P. (1995). The importance of an early emphasis on L2 vocabulary. *LANGUAGE TEACHER-KYOTO-JALT-*, *19*, 8–11. Retrieved from http://www.lognostics.co.uk/vlibrary/meara1995f.pdf

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meara, P. (2005). Lexical Frequency Profiles: A Monte Carlo Analysis. *Applied Linguistics*, *26*(1), 32–47. doi:10.1093/applin/amh037

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*(2), 142–154.

Meara, P., & Milton, J. (2003). The Swansea Levels Test. Newbury: Express.

Meara, P., & Milton, J. (2006). X-Lex: the Swansea Vocabulary Levels Test. In C. Coombe, P. Davidson, & L. D (Eds.), *the 7th and 8th Current Trends in English Language testing (CTELT) Conference* (pp. 29–39). Dubai: TESOL Arabia.

Mehrpour, S., & Rahimi, M. (2010). The impact of general and specific vocabulary knowledge on reading and listening comprehension: A case of Iranian EFL learners. *System*, *38*(2), 292–300.

Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In J. Daller, H., Milton, J. & Treffers-Daller (Ed.), *Modelling and assessing vocabulary knowledge* (pp. 47–58). Cambridge: Cambridge University Press. Retrieved from http://www.cup.es/servlet/file/store7/item5633259/version1/FT Vocab_Book chapter_Modelling and Assessing-2-Desc.pdf

Milton, J. (2009). *Measuring second language vocabulary acquisition*. (D. Singleton, Ed.)*Exposure* (Vol. 45, p. 287). Multilingual Matters.

Milton, J., & Hopkins, N. (2005). Aural Lex. Swansea.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. *Further Insights into*, *1*(11), 83–97. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract

Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*(1), 73–98. doi:10.1191/0265532206lt321oa

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System*, *32*(1), 75–87. doi:10.1016/j.system.2003.05.001

Nagy, W. (1997). In The Role of Context in First and Second Language Learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 64–83). Cambridge University Press.

Nation, P. (1982). Beginning to learn foreign vocabulary: A review of the research. *RELC Journal*, *13*(1), 14–36.

Nation, P. (1990). *Teaching and Learning Vocabulary*. *Teaching and learning vocabulary: Bringing research to practice* (p. 1). Lawrence Erlbaum.

Nation, P. (2001). *Learning vocabulary in another language*. (M. H. Long & J. C. Richards, Eds.)*Victoria University of Wellington English Language Institute Occasional Publication No 19* (p. 1). Cambridge University Press.

Nation, P. (2005). Teaching Vocabulary. *Asian EFL Journal*, *7*(3).

Nation, P. (2006). How Large a Vocabulary Is Needed for Reading and Listening? *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–81.

Nation, P. (2007a). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 35–43). Cambridge: Cambridge University Press.

Nation, P. (2007b). The Four Strands. *Innovation in Language Learning and Teaching*, *1*(1), 2–13.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13. Retrieved from http://www.victoria.ac.nz/lals/staff/Publications/paul-nation/2007-Beglar-TLT.pdf

Nation, P., & Gu, P. (2007). *Focus on Vocabulary*. *Focus*. Sydney: NCELTR, Macquarie University.

Nation, P., & Meara, P. (2010). Vocabulary. In N. Schmitt (Ed.), *An Introduction to Applied Linguistics* (second edi., pp. 34–52). London: Hodder Education.

Nation, P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*(2), 355–380.

Nikitina, L., & Furuoka, F. (2006). Re-examining Horwitz ' s Beliefs About Language Learning Inventory ( BALLI ) in the Malaysian Context. *Electronic Journal of Foreign Language Teaching*, *3*(2), 209–219. Retrieved from http://e-flt.nus.edu.sg/v3n22006/nikitina.htm

Nordstokke, D., & Zumbo, B. (2007). A Cautionary Tale About Levene's Tests for Equal Variances. *Journal of Educational Research & Policy …*, 1–14. Retrieved from http://educ.ubc.ca/faculty/zumbo/papers/Nordstokke_Zumbo_JERPS_2007_reprint.pdf

O'Sullivan, B., & Weir, C. J. (2011). Test Development and Validation. In Palgrave Advances in Linguistics (Ed.), *Language Testing: Theories and Practices* (pp. 13–32). Basingstoke, UK: Palgrave Macmillan.

Oxford, R., & Crookall, D. (1990). Vocabulary Learning : A Critical Analysis of Techniques. *TESL Canada Journal*, *7*(2), 9–30.

Paribakht, T. S., & Wesche, M. (1999). Reading and "Incidental" L2 Vocabulary Acquisition: An Introspective Study of Lexical Inferencing. *Studies in Second Language Acquisition*, *21*(2), 195–224.

Pellicer-Sanchez, a., & Schmitt, N. (2012). Scoring Yes-No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, *29*(4), 489–509. Retrieved from http://ltj.sagepub.com/cgi/doi/10.1177/0265532212438053

Perkins, K., Brutten, S., & Pohlmann, J. (1989). First and second language reading comprehension. *RELC Journa*, *10*(2), 1–9.

Pichette, F., Segalowitz, N., & Connors, K. (2003). Impact of maintaining L1 reading skills on L2 reading skill development in adults: Evidence from speakers of Serbo-Croatian learning French. *The Modern Language Journal*, *87*, 391–403.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1–28.

Pretorius, E. (2000). " What they can't read will hurt them": reading and academic achievement. *INNOVATION-PIETERMARITZBURG-*, *21*, 33–41. Retrieved from http://www.innovation.ukzn.ac.za/InnovationPdfs/No21pp33-41Pretorius.pdf

Punch, K. (2005). *Introduction to social research: Quantitative and qualitative approaches* (Vol. 311). SAGE Publications Ltd.

Qian, D. (1996). ESL Vocabulary Acquisition: Contextualization and Decontextualization. *Canadian Modern Language Review*, *53*(1), 120–42.

Qian, D. (1999). Assessing the Roles of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension. *Canadian Modern Language Review La Revue Canadienne Des Langues Vivantes*, *56*(2), 282–308.

Qian, D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning*, *52*(3), 513–536.

Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, *21*(1), 28–52.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*(3), 355–371.

Read, J. (2000). *Assessing vocabulary*. *Applied Linguistics*. Cambridge: Cambridge University Press.

Read, J. (2004). 7. Research in Teaching Vocabulary. *Annual Review of Applied Linguistics*, *24*, 146–161.

Read, J. (2007). Second Language Vocabulary Assessment: Current Practices and New Directions. *IJES*, *7*(2), 105–125.

Richard, J. P. J. (2011). Does size matter? The relationship between vocabulary breadth and depth. *Sophia Interntional Review*, *33*, 107. Retrieved from http://fla-sir.weebly.com/uploads/2/4/7/1/2471121/sir33-richard.pdf

Richards, J. C. (1976). The Role of Vocabulary Teaching. *Tesol Quarterly*, *10*(1), 77–89.

Richmond, K., & Zimmerman, C. (2007). *Inside Reading 4 Instructor Pack: The Academic Word List in Context*. Oxford University Press.

Schmitt, D., & Schmitt, N. (2005). *Focus on Vocabulary: Mastering the Academic Word List*. New York: Pearson Education.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Schmitt, N. (2008a). *Review article: Instructed second language vocabulary learning*. *Language Teaching Research* (Vol. 12, pp. 329–363).

Schmitt, N. (2008b). Teaching Vocabulary. *Pearson Education*, (2006).

Schmitt, N. (2010a). Key issues in teaching and learning vocabulary. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Insights into Non-native Vocabulary Teaching and Learning*. Multilingual Matters.

Schmitt, N. (2010b). *Researching Vocabulary: A Vocabulary Research Manual*. *Studies in Second Language Acquisition* (Vol. 19). Palgrave Macmillan.

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: Theoretical underpinnings and practical suggestions. *ELT Journal*, *49*(2), 133–143.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Shefelbine, J. (1990). Student factors related to variability in learning word meanings from context. *Journal of Literacy Research*, *22*(1), 71–97. doi:10.1080/10862969009547695

Siebert, L. (2003). Student and teacher beliefs about language learning. *The ORTESOL Journal*, *21*, 7–39.

Sokmen, A. (1997). Current Trends in Teaching Second Language Vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary Description acquisition and Pedagogy* (pp. 237–257). Cambridge: Cambridge University Press.

Sonbul, S., & Schmitt, N. (2009). Direct teaching of vocabulary after reading: is it worth the effort? *ELT Journal*, *64*(3), 253–260.

Srichamnong, N. (2008). Incidental EFL Vocabulary Learning : The Effects of Interactive Multiple-Choice Glosses. *Design*, (Belisle 1997).

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139–152.

Stæhr, L. S. (2009). Vocabulary Knowledge and Advanced Listening Comprehension in English As a Foreign Language. *Studies in Second Language Acquisition*, *31*(04), 577.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–253). Newbury House.

Tannenbaum, K., Torgesen, J., & Wagner, R. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, *10*(4), 381–398.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied PsychoLinguistics*, *22*(2), 217–234.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader ? *Reading in a Foreign Language*, *15*(2).

Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics*, *28*(1), 46–65.

Webb, S. (2008). Receptive and Productive Vocabulary Sizes of L2 Learners. *Studies in Second Language Acquisition*, *30*(01), 79–95.

Webb, S. (2009). The Effects of Pre-learning Vocabulary on Reading Comprehension and Writing. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *65*(3), 441–470.

Webb, S. (2010). Pre-learning low-frequency vocabulary in second language television programmes. *Language Teaching Research*, *14*(4), 501–515.

Weir, C. J. C. (2005). *Language testing and validation: An evidence-based approach*. (C. N. Candlin, Ed.) (Vol. 9, p. 4039). Basingstoke, UK: Palgrave Macmillan.

West, M. (1953). *A General Service List of English Words*. London: Longman, Green and Co.

Wilkins, D. (1972). *Linguistics in Language Teaching*. Cambridge, MA: MIT Press.

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication*, *3*(2), 215–229.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, (57).

Zimmerman, C. (1997a). Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly*, *31*(1), 121–140.

Zimmerman, C. (1997b). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 5–19). New York: Cambridge University Press.

# APPENDICES

# Appendix A

*English XK-Lex Vocabulary Test 1*

Please look at these words. Some of these words are real English words and some are not but are made to look like real words. Please tick the words that you know or can use. Here is an example.

Version: **A** **cat** ✓

**EVST Score:**

**Your student number**:

**Thank you for your help.**

| | | | | |
|---|---|---|---|---|
| **New** | **Commerce** | **Organise** | **Accuse** | **Victory** |
| **Gummer** | **Tindle** | **Wookey** | **Candish** | **Skave** |
| **Word** | **Dust** | **Fountain** | **Tend** | **Jewel** |
| **Near** | **Nonsense** | **Movement** | **Landing** | **Reliable** |
| **Peace** | **Fond** | **Likely** | **Volume** | **Harden** |
| **Produce** | **Sweat** | **Provide** | **Tube** | **Sorrow** |
| **You** | **Cap** | **Castle** | **Liner** | **Dial** |
| **Wife** | **Worry** | **Steam** | **Previous** | **Enclose** |
| **Do** | **Plenty** | **Steady** | **Style** | **Sneeze** |
| **Add** | **Guide** | **Pole** | **Outline** | **Apparatus** |
| **Kilp** | **Broy** | **Orrade** | **Plaudate** | **Overend** |
| **Build** | **Pump** | **Guest** | **Keeper** | **Roast** |
| | | | | |
| **Prosecutor** | **Addict** | **Gulp** | **Idleness** | **Carnation** |
| **Samphirate** | **Treadway** | **Darch** | **callisthemia** | **Mordue** |
| **Referral** | **Detachment** | **Thud** | **Blizzard** | **Plaintively** |
| **Illuminate** | **Unsure** | **Assassin** | **Rut** | **Gurgle** |
| **Gown** | **Reinforcement** | **Wrench** | **Incessant** | **Heal** |
| **Verge** | **Enlightenment** | **Backdrop** | **Blunder** | **go-between** |
| **Counsellor** | **Workman** | **Unfold** | **springboard** | **common-law** |
| **Skipper** | **Feudal** | **Upheaval** | **Shrapnel** | **Locket** |
| **Authorise** | **Quartet** | **Animation** | **Skip** | **Nudge** |
| **Sour** | **Psychic** | **Banish** | **Bastion** | **Anger** |
| **Neminary** | **Fallity** | **Treggle** | **Snape** | **Tearle** |
| **Holly** | **Appropriation** | **Peninsula** | **Maroon** | **Contrive** |
| | | | | |

# Appendix B

*English XK-Lex Vocabulary Test 2*

Please look at these words. Some of these words are real English words and some are not but are made to look like real words. Please tick the words that you know or can use. Here is an example.

Version: **B** **cat** ✓

**EVST Score:**

**Your student number**:

**Thank you for your help.**

| | | | | |
|---|---|---|---|---|
| **make** | **Advice** | **Generous** | **Cure** | **Victory** |
| **Anand** | **Trudgeon** | **Snell** | **hammond** | **arbus** |
| **Turn** | **perform** | **Rabbit** | **Pat** | **Opponent** |
| **Doubt** | **Luck** | **Cough** | **Court** | **Feast** |
| **Start** | **Fierce** | **Sense** | **reaction** | **item** |
| **Ready** | **Strict** | **Announce** | **workshop** | **fortune** |
| **Person** | **Collar** | **Prepare** | **leadership** | **simplicity** |
| **open** | **wire** | **Drag** | **reference** | **overlook** |
| **Fact** | **Comfort** | **Sight** | **emphasise** | **scorn** |
| **Sure** | **Discipline** | **Situation** | **seed** | **respect** |
| **Widgery** | **Inertible** | **Loring** | **craddock** | **encopulate** |
| **Write** | **Pour** | **Dive** | **calculate** | **junction** |
| | | | | |
| **Dependency** | **Convergence** | **Cape** | **tireless** | **cylinder** |
| **Chibberv** | **Fallology** | **Atone** | **lebrucious** | **outpanner** |
| **Descendant** | **Alley** | **Conscientious** | **eloquence** | **allure** |
| **Playground** | **Cutter** | **paw** | **spurt** | **atone** |
| **Attachment** | **Consultative** | **Reap** | **recoup** | **ruby** |
| **Hurdle** | **Contamination** | **Extremist** | **buoyancy** | **dicey** |
| **Offering** | **Hierarchical** | **Adorn** | **squeak** | **coterie** |
| **Denote** | **cram** | **Rejoin** | **sighting** | **conundrum** |
| **Accumulation** | **rivalry** | **Admirer** | **Stout** | **chipboard** |
| **Simplify** | **shark** | **Animated** | **Braid** | **barn** |
| **Proom** | **Skave** | **Spalding** | **Coath** | **charlett** |
| **Binary** | **Severity** | **Questionable** | **Breed** | **maggot** |
| | | | | |

# Appendix C

| | | | |
|---|---|---|---|
| 1 area | | 1 alter | |
| 2 contract | _____ written agreement | 2 coincide | _____ change |
| 3 definition | _____ way of doing something | 3 deny | _____ say something is |
| 4 evidence | _____ reason for believing | 4 devote | not true |
| 5 method | something is or | 5 release | _____ describe clearly |
| 6 role | is not true | 6 specify | and exactly |

| | | | |
|---|---|---|---|
| 1 debate | | 1 correspond | |
| 2 exposure | _____ plan | 2 diminish | _____ keep |
| 3 integration | _____ choice | 3 emerge | _____ match or be in |
| 4 option | _____ joining something | 4 highlight | agreement with |
| 5 scheme | into a whole | 5 invoke | _____ give special |
| 6 stability | | 6 retain | attention to something |

| | | | |
|---|---|---|---|
| 1 access | | 1 bond | |
| 2 gender | _____ male or female | 2 channel | _____ make smaller |
| 3 implementation | _____ study of the mind | 3 estimate | _____ guess the number |
| 4 license | _____ entrance or way in | 4 identify | or size of something |
| 5 orientation | | 5 mediate | _____ recognizing and |
| 6 psychology | | 6 minimize | naming a person or thing |

| | | | |
|---|---|---|---|
| 1 accumulation | | 1 explicit | |
| 2 edition | _____ collecting things over time | 2 final | _____ last |
| 3 guarantee | _____ promise to repair a | 3 negative | _____ stiff |
| 4 media | broken product | 4 professional | _____ meaning `no' or `not' |
| 5 motivation | _____ feeling a strong reason | 5 rigid | |
| 6 phenomenon | or need to do something | 6 sole | |

| | | | |
|---|---|---|---|
| 1 adult | | 1 abstract | |
| 2 exploitation | _____ end | 2 adjacent | _____ next to |
| 3 infrastructure | _____ machine used to move | 3 controversial | _____ added to |
| 4 schedule | people or goods | 4 global | _____ concerning the |
| 5 termination | _____ list of things to do at | 5 neutral | whole world |
| 6 vehicle | certain times | 6 supplementary | |

*Version A of the Academic word list test (receptive)*

# Appendix D

*Version B of AWL TEST*

| | |
|---|---|
| 1. benefit | |
| 2. labour | ___ work. |
| 3. percent | ___ part of 100 . |
| 4. principle | ___ general idea used to guide one's actions. |
| 5. source | |
| 6. survey | |

| | |
|---|---|
| 1. achieve | |
| 2. conceive | ___ change. |
| 3. grant | ___ connect together. |
| 4. link | ___ finish successfully. |
| 5. modify | |
| 6. offset | |

| | |
|---|---|
| 1. element | |
| 2. fund | ___ money for a special. Purpose |
| 3. layer | |
| 4. philosophy | ___ skilled way of doing Something |
| 5. proportion | |
| 6. technique | ___ study of the meaning of life |

| | |
|---|---|
| 1. convert | |
| 2. design | ___ keep out. |
| 3. exclude | ___ stay alive. |
| 4. facilitate | ___ change from one thing into another |
| 5. indicate | |
| 6. survive | |

| | |
|---|---|
| 1. consent | |
| 2. enforcement | ___ total. |
| 3. investigation | ___ agreement or permission. |
| 4. parameter | ___ trying to find information about something. |
| 5. sum | |
| 6. trend | |

| | |
|---|---|
| 1. anticipate | |
| 2. compile | ___ control something skilfully. |
| 3. convince | ___ expect something will. Happen |
| 4. denote | |
| 5. manipulate | ___ produce books and Newspapers |
| 6. publish | |

| | |
|---|---|
| 1. decade | |
| 2. fee | ___ 10 years. |
| 3. file | ___ subject of a discussion. |
| 4. incidence | ___ money paid for services. |
| 5. perspective | |
| 6. topic | |

| | |
|---|---|
| 1. equivalent | |
| 2. financial | ___ most important. |
| 3. forthcoming | ___ concerning sight. |
| 4. primary | ___ concerning money. |
| 5. random | |
| 6. visual | |

| | |
|---|---|
| 1. colleague | |
| 2. erosion | ___ action against the law. |
| 3. format | ___ wearing away gradually. |
| 4. inclination | ___ shape or size of something |
| 5. panel | |
| 6. violation | |

| | |
|---|---|
| 1. alternative | |
| 2. ambiguous | ___ last or most important. |
| 3. empirical | ___ something different that can be chosen. |
| 4. ethnic | |
| 5. mutual | ___ concerning people from a certain nation. |
| 6. ultimate | |

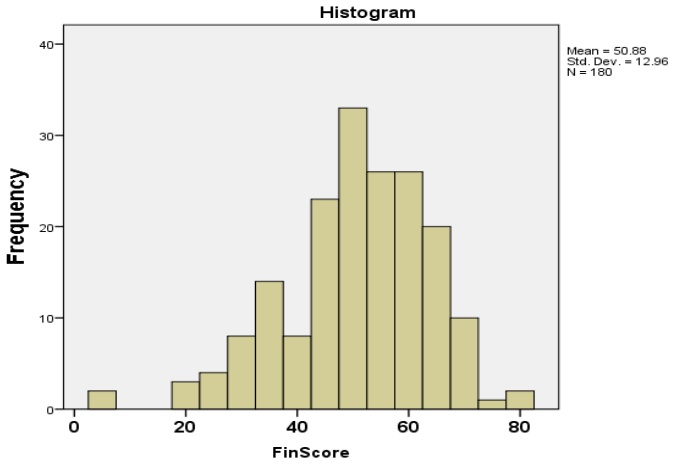
## ***Appendix E*** *Level - AWL (productive Academic Word List)*

For each given word/phrase, provide another word with the same meaning. The word should begin with the letter provided

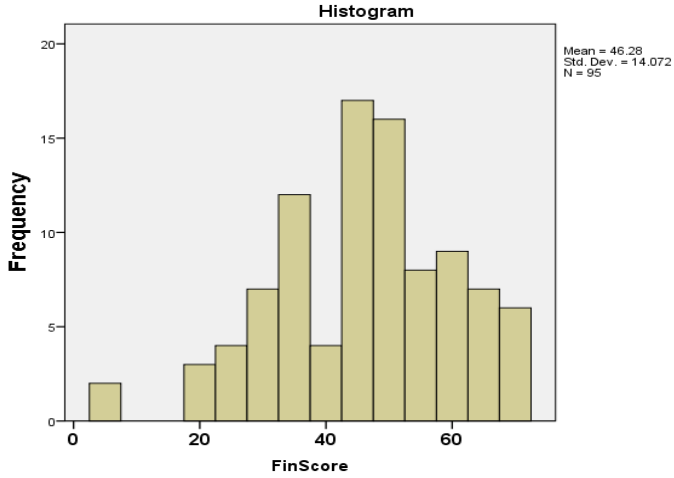
| | |
|---|---|
| 1. general idea used to guide one's actions | p_____ |
| 2. agreement or permission | c_____ |
| 3. moving from one place to another | m_____ |
| 4. 10 years | d_____ |
| 5. acting against the law | v_____ |
| 6. wearing away gradually | e_____ |
| 7. something different that can be chosen | a_____ |
| 8. match or be in agreement with | c_____ |
| 9. give special attention to something | h_____ |
| 10. keep | r_____ |
| 11. control something skillfully | m_____ |
| 12. expect something will happen | a_____ |
| 13. last or most important | u_____ |
| 14. joining something into a whole | i_____ |
| 15. choice | o_____ |
| 16. plan | s_____ |
| 17. entrance or way in | a_____ |
| 18. collecting things over time | a_____ |
| 19. list of things to do at certain times | s_____ |
| 20. end | t_____ |
| 21.to change | a_____ |
| 22. describe clearly and exactly | s_____ |
| 23. change from one thing into another | c_____ |
| 24. guess the number or size of something | e_____ |
| 25. recognize and name a person or a thing | i_____ |
| 26. stiff | r_____ |
| 27. most important and noticeable | p_____ |
| 28. happening after | s_____ |
| 29. very near | a_____ |
| 30. added to | s_____ |

# Appendix F

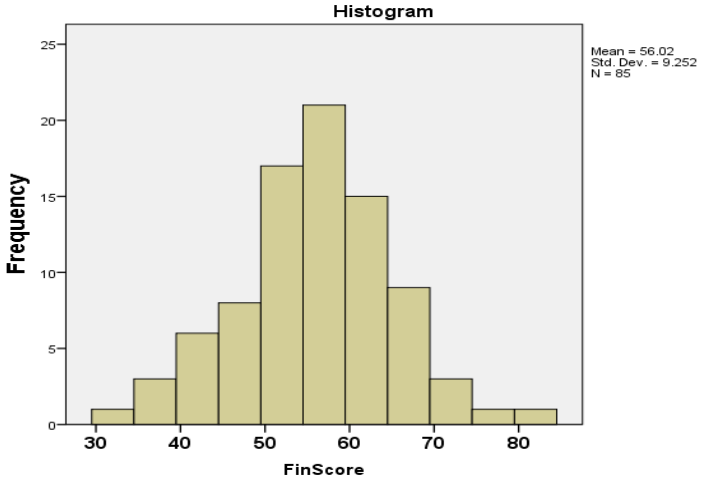
*Histogram graph for XK-lex data for RUni1and SUni2*




*Histogram graph for XK-lex data for SUni1*




*Histogram graph for XK-lex data for RUni1*



# Appendix G

*The entire AWL words and number of types produced by each writer: Data from Chapter 4*

| | | | repeat |
|---|---|---|---|
| 103 | Pre | cultural_[1] economic_[3] emphasis_[1] **factor**_[2] finally_[1] generation_[1] job_[1] media_[1] source_[1] theory | 1 |
| | Post | benefit_[3] cited_[1] conclusion_[1] contrast_[1] create_[3] creative_[1] ensure_[1] establish_[1] evaluate_[1] **factors**_[2] finance_[2] furthermore_[1] goals_[1] impact_[1] indicate_[1] indicated_[1] individual_[2] investment_[2] network_[1] nevertheless_[2] process_[3] stress_[1] team_[1] technical_[1] welfare_[1] whereas_[1] | |
| 105 | Pre | adult_[1] **conclusion**_[1] culture_[1] **instance**_[1] job_[1] option_[1] significant_[1] | 2 |
| | Post | beneficial_[1] benefits_[1] channels_[1] cited_[1] **conclusion**_[1] creative_[2] despite_[1] environment_[1] establish_[2] expand_[1] **instance**_[1] invest_[1] network_[1] networks_[1] sufficient_[1 | |
| 113 | Pre | computers_[1] **creativity**_[1] **job**_[1] methods_[1] | 2 |
| | Post | benefits_[1] conclusion_[1] create_[1] **creativity**_[2] defined_[1] finally_[1] financial_[1] individual_[1] **job**_[1] obviously_[1] policies_[1] transfer_[1] welfare_[1] | |
| 114 | Pre | academic_[2] computer_[1] conclusion_[1] decade_[1] economic_[1] **instance**_[1] **job**_[1] major_[2] medical_[1] normal_[1] topic_[1] tradition_[1] | 2 |
| | Post | benefit_[1] challenge_[1] cited_[1] conflict_[1] consumer_[1] hence_[3] initially_[1] **instance**_[2] **job**_[1] obviously_[1] partner_[2] sum_[1] | |
| 122 | Pre | **achieve**_[1] acknowledge_[1] **conclusion**_[1] cultural_[1] **establish**_[1] institutions_[1] invest_[1] **jobs**_[1] **requirement**_[1] | 5 |
| | Post | **achieved**_[1] appropriate_[1] **conclusion**_[1] consequence_[1] creative_[1] creativity_[1] elements_[1] **establish**_[2] exposure_[1] factors_[1] financial_[1] fundamental_[1] globalization_[2] highlighted_[1] identified_[1] illustration_[1] indicated_[1] initiate_[1] **job**_[2] labor_[1] location_[1] logical_[1] negative_[1] notion_[1] obvious_[1] occupation_[1] positive_[1] range_[1] relevant_[1] **requirements**_[1] sector_[1] specific_[1] traditional_[1] | |
| M e e 129 | Pre | complex_[1] dramatically_[1] environment_[1] feature_[2] furthermore_[1] indication_[1] **job**_[1] **major**_[2] obvious_[1] primary_[1] **significantly**_[1] specific_[1] | 3 |
| | Post | communicate_[1] contrast_[1] contribute_[2] creative_[1] enable_[1] evaluate_[1] financial_[1] flexible_[1] ignored_[1] illustrate_[1] individuals_[1] inevitably_[1] insecurity_[1] investment_[1] **job**_[2] **major**_[1] regulations_[1] rejected_[1] restriction_[1] role_[1] schedule_[2] scheme_[1] **significant**_[1] stable_[1] | |
| t i n g 138 2 | Pre | academic_[1] academics_[2] acquire_[1] area_[1] areas_[1] collapsing_[1] economies_[1] economy_[3] **furthermore**_[1] global_[1] globalization_[1] **hence**_[1] job_[1] relied_[1] rely_[1] require_[1] requirements_[1] significantly_[1] specific_[1] survive_[2] technology_[4] undeniable_[2] | 2 |
| | Post | accumulation_[1] analysed_[1] benefit_[1] benefits_[4] consistently_[1] contrast_[1] contributes_[1] creative_[1] dramatically_[1] **furthermore**_[1] **hence**_[1] inadequate_[1] inclined_[1] jobs_[1] perceive_[1] periodically_[1] promoted_[2] promotion_[1] selecting_[1] strategies_[1] strategy_[1] stress_[1] sum_[1] technical_[1] transferable_[1] whereas_[1] | |
| 4 J a 144 | Pre | conduct_[1] focus_[1] grades_[1] **job**_[3] motivate_[1] primary_[2] sum_[1] text_[1] trend_[1] | 1 |
| | Post | aids_[1] cited_[1] conception_[1] conclusion_[1] contrast_[1] factor_[1] financial_[1] goals_[1] implies_[1] incomes_[1] investment_[1] issue_[1] **job**_[2] network_[1] regulation_[1] restriction_[3] secure_[1] security_[1] similar_[1] stable_[2] volumes_[1] | |

# Appendix H

*The entire AWL words and number of types produced by each writer: Data from Chapter 5*

| | | | repeat |
|---|---|---|---|
| 315 | Pre | benefit_[1] community_[2] environment_[2] normal_[1] policy_[1] stress_[1] | |
| | Post | 0 | |
| 320 | Pre | environment_[2] final_[1] generation_[1] prohibit_[1] | |
| | Post | 0 | |
| 306 | Pre | appropriate_[1] computer_[1] technology_[1] | |
| | Post | environment_[3] environmental_[4] sources_[1] | |
| 301 | Pre | communicate_[1] link_[1] task_[2] | |
| | Post | consequence_[1] consume_[1] environment_[3] environmental_[1] generation_[1] global_[1] hence_[1] method_[1] minimum_[1] summary_[1] transition_[1] | |
| 210 | Pre | area_[1] chemical_[2] conclude_[1] environment_[3] globe_[1] issue_[1] respond_[1] source_[1] technical_[1] vary_[1] | |
| | Post | classic_[1] contrast_[1] definite_[1] energy_[1] intelligence_[1] relax_[1] unique_[1] | |
| 222 | Pre | affect_[3] alternative_[1] consequent_[1] cycle_[1] environment_[4] occur_[1] random_[2] rely_[1] specific_[1] sum_[1] | |
| | Post | area_[2] conclude_[1] relax_[2] technology_[1] | |
| 232 | Pre | environment_[2] | |
| | Post | relax_[3] | |
| 234 | Pre | conclude_[1] cooperate_[1] environment_[7] globe_[1] individual_[1] | |
| | Post | conclusion_[1] finally_[1] located_[1] | 1 |

## Appendix I:

*Learners' errors in writing corrected before the analysis:*

**Data from the preliminary study:**

| ID | passage | paragraph | Line | Word | What do you think means |
|---|---|---|---|---|---|
| 122 | 2 | 4 | 1 | *fundaments | Fundamental |
| 122 | 2 | 4 | 5 | *fasters | Factors |
| 122 | 2 | 5 | 2 | *hililted | Highlighted |
| 122 | 1 | 1 | 6 | *achive | Achieve |
| 122 | 1 | 2 | 3 | *aknaledge | Acknowledge |
| 103 | 1 | 1 | 1 | *Globalize | Globalization |
| 129 | 2 | 4 | 4 | *investmented | investment |
| 114 | 1 | 1 | 1 | *decide | decade |
| 138 | 2 | 3 | 9 | *an adequente | an adequate |
| 138 | 1 | 2 | 1 | *cotlapsion | collapsing |

**Data from the main study:**

| ID | passage | paragraph | Line | Word | What do you think means |
|---|---|---|---|---|---|
| 232 | 2 | 1 | 1 | *relaxe | relax |
| 234 | 1 | 2 | 4 | *defcte | Defeat |
| 232 | 1 | 2 | 2 | *=factors | factories |

## **Appendix J:** Learners' original questionnaire

**1-Basic Information**

1. **Your class/teacher group:** _____

2. **How many years have you been studying English:**

A. 0—5 years B. 6—7 years C. 8—9 years D. 10—11 years E. over 12 years

3. **Have you taken any of these tests lately and what was your score**:

IELTS _____ TOEFL ◯ _____ SWELT ◯ _____ other please specify: _____

4. **After your EAP study, you are going to take:**

A. undergraduate programme B. postgraduate programme (Masters level) C. research degree (Dr level) D. none of these.

**2. Your thoughts about** *Vocabulary Learning* **that you think occurred during this term:**

| **1 means** *Totally disagree* and **5 means** *Totally agree* | | | | | |
|---|---|---|---|---|---|
| 5. In general, I think I have learned a lot of academic vocabulary during this course. | **1** | **2** | **3** | **4** | **5** |
| 6. I think I have learned a lot of academic vocabulary during this course and I'm able to use this in writing and speaking | **1** | **2** | **3** | **4** | **5** |
| 7- There was a specific focus on the academic vocabulary during the course. | **1** | **2** | **3** | **4** | **5** |
| 8. The learning of the academic vocabulary was enhanced during study skills (Reading, Writing, Listening etc) in this course. | **1** | **2** | **3** | **4** | **5** |
| 9. The academic vocabulary is important for my language study. | **1** | **2** | **3** | **4** | **5** |
| 10. The academic vocabulary is important for my further studies. | **1** | **2** | **3** | **4** | **5** |
| 11. I have been taught some academic vocabulary and I think it is enough for my further studies and I do not need more. | **1** | **2** | **3** | **4** | **5** |
| 12. I value the vocabulary that I have learned but I wished I had more during the course. | **1** | **2** | **3** | **4** | **5** |
| 13. I regularly studied vocabulary outside class using notes or guidelines set out by my teacher | **1** | **2** | **3** | **4** | **5** |
| 14. I regularly studied vocabulary outside class on my own initiative. | **1** | **2** | **3** | **4** | **5** |

**15. Did you use any textbook which focuses on vocabulary or academic vocabulary during the course?**

A: Yes ◯ (title if you remember): ..... B: No ◯

**16. How would you describe your vocabulary learning this Summer? (**more than one option is possible**)**

A) In class aided by teacher. B) Homework aided by teacher. C) Self vocabulary learning. D) None.

**Your name Please:** .....

## Second questionnaire after minor amendment

**1. Basic Information (Level 5)**

1. **Your class/teacher group:** _____

2. **How many years have you been studying English:**

A. 0—5 years B. 6—7 years C. 8—9 years D. 10—11 years E. over 12 years

3. **Have you taken any of these tests in the last two years, and what was your score**?

IELTS ◯ _____ TOEFL ◯ _____ other please specify: _____

4. **Which major have you been offered to study at King Saud University?**

A. Science and Engineering. C. Humanities D. Medicine

**2. Your thoughts about *Vocabulary Learning* that you think occurred during this term:**

| | Disagree | | | | Agree |
|---|---|---|---|---|---|
| 5. In general, I think I have learned a lot of academic vocabulary during this course. | **1** | **2** | **3** | **4** | **5** |
| 6. I think I have learned a lot of academic vocabulary during this course and I'm able to use this in writing and speaking | **1** | **2** | **3** | **4** | **5** |
| 7- There was a specific focus on the academic vocabulary during the course. | **1** | **2** | **3** | **4** | **5** |
| 8. The learning of the academic vocabulary was enhanced during study skills (Reading, Writing, Listening etc) in this course. | **1** | **2** | **3** | **4** | **5** |
| 9. The academic vocabulary is important for my language study. | **1** | **2** | **3** | **4** | **5** |
| 10. The academic vocabulary is important for my further studies. | **1** | **2** | **3** | **4** | **5** |
| 11. I have been taught some academic vocabulary in class. | **1** | **2** | **3** | **4** | **5** |
| 12. The academic vocabulary I now know is enough for my further studies. | **1** | **2** | **3** | **4** | **5** |
| 13. I value the vocabulary that I have learned but I wished I had more during the course. | **1** | **2** | **3** | **4** | **5** |
| 14. I regularly studied vocabulary outside class using notes or guidelines set out by my teacher | **1** | **2** | **3** | **4** | **5** |
| 15. I regularly studied vocabulary outside class on my own initiative. | **1** | **2** | **3** | **4** | **5** |

**16. Did you use any textbook which focuses on vocabulary or academic vocabulary during the course?**

A: Yes ◯ (title if you remember): ..... B: No ◯

**17. How would you describe your vocabulary learning during this term? (**more than one option is possible**)**

A) In class aided by teacher. B) Homework set by teacher. C) Self vocabulary learning. D) None.

**Your name Please:** .....

**Origial Questionnaire** adopted from (Alothman, 2009).

1. **Your age:** A. under 20 B. between 20—25 C. over 25

2. **Your sex:** A. male B. female

3. **You are from** _____ (Country)

4. **Your group:** A. Red B. Blue C. Green D. Orange E. Purple.

5. **How many years have you been studying English:**

A. 0—5 years B. 6—7 years C. 8—9 years D. 10—11 years E. over 12 years

6. **Have you taken IELTS lately**,_____

If yes, please say when?_____, and what was your overall score: _____

7. **After your EAP study, you are going to take:**

A. undergraduate programme B. postgraduate programme C. neither.

**2.Your thoughts about the text book** *Focus On Vocabulary* **that you have been studying this term:**

| **1** *Totally agree* **5** *Totally disagree* | | | | | |
|---|---|---|---|---|---|
| 8. In general, I have befitted from the text book *Focus On Vocabulary* | **1** | **2** | **3** | **4** | **5** |
| 9. The vocabulary that I have learned has influenced my Reading | **1** | **2** | **3** | **4** | **5** |
| 10. I have recognized the vocabulary that I have learned in my Listening | **1** | **2** | **3** | **4** | **5** |
| 11. I have used some vocabulary that I have learned in my Writing | **1** | **2** | **3** | **4** | **5** |
| 12. I have used some vocabulary that I have learned in my Speaking | **1** | **2** | **3** | **4** | **5** |
| 13. I value the book because of the selection of the words. | **1** | **2** | **3** | **4** | **5** |
| 14. I value the book because it gives many forms of each word. | **1** | **2** | **3** | **4** | **5** |
| 15. I value the book as it teaches me the right use of words. | **1** | **2** | **3** | **4** | **5** |
| 16. The vocabulary that I have learned is enough for my further studies and I do not need more. | **1** | **2** | **3** | **4** | **5** |
| 17. I value the vocabulary that I have learned and I wished we had more | **1** | **2** | **3** | **4** | **5** |
| 18. I did not like the book and I wished we have studied something else. | **1** | **2** | **3** | **4** | **5** |
| 19. The book was very difficult. | **1** | **2** | **3** | **4** | **5** |
| 20. I now realize the importance of academic vocabulary for my study. | **1** | **2** | **3** | **4** | **5** |
| 21. I often read the book outside the class | **1** | **2** | **3** | **4** | **5** |
| 22. I will study the book after the term finishes. | **1** | **2** | **3** | **4** | **5** |
| 23. I recommend this book for future students. | **1** | **2** | **3** | **4** | **5** |

**24. Do you focus on vocabulary in your studying?** A: Yes B: No

**25. Have read any book which focuses on vocabulary before?** A: Yes title:……B: No

**26. How important is academic vocabulary for your study at the EAS course?**

……

## Teachers' open-ended survey

*This questionnaire evaluates the learning of academic vocabulary in the PY program English courses.*

Please mark the response that most closely matches your opinion.

**1. PY program should have a course that focuses specifically on academic vocabulary.**

o Strongly agree o Agree o No opinion o Disagree o Strongly Disagree

**2. Learning academic vocabulary within the course curriculum (language skills and input) is enough.**

o Strongly agree o Agree o No opinion o Disagree o Strongly Disagree

**3. Instead of allocating some of the schooling time to teach vocabulary, it is better to enhance more important language skills (for example writing).**

o Strongly agree o Agree o No opinion o Disagree o Strongly Disagree

**4. How would you describe the learning of Academic vocabulary (specifically the Academic Word List or AWL) that took place in your class? (more than one option is possible)**

A) I covered some of it in class time ☐
B) I covered most or all of it in class time ☐
C) I encouraged students to cover most or all of it in their time ☐
D) I was not aware of such a list for this programme ☐
E) There was a list but I decided not work on it ☐

**5. How motivated were your students to learn academic vocabulary**

o Highly o Somewhat o Not sure o Little o Not at all (Please explain your response below)

……

6. If you have covered parts of the AWL list, about how many words have you covered and what do you think is feasible within one term?

……
……

7. To what extent you think the words students have learned **have influenced their other specific language skills**

| | | | | | |
|---|---|---|---|---|---|
| **Reading** | o Strongly | o Somewhat | o Not sure | o Little | o Not at all |
| **Writing** | o Strongly | o Somewhat | o Not sure | o Little | o Not at all |
| **Speaking** | o Strongly | o Somewhat | o Not sure | o Little | o Not at all |
| **Listening** | o Strongly | o Somewhat | o Not sure | o Little | o Not at all |

**Please, write the name of the group/s that you taught: .....**

## **Appendix K:** Multivariate Tests of effects of variables: *major of study, number of years of EFL education, or proficiency* and *the level of study*

### Major of study effect:

| | | Value Label | N |
|---|---|---|---|
| Major | 1 | SciAndEng | 1 |
| | 2 | Humanities | 58 |
| | 3 | Medcine | 18 |

| Dependent Variable | Df | M-Square | F | Sig. |
|---|---|---|---|---|
| **Rates**: | | | | |
| Learned AWL | 2 | .219 | .210 | .811 |
| Learn And Can Write Speak AWL | 2 | .068 | .067 | .936 |
| Course Focused On AWL | 2 | .147 | .119 | .888 |
| AWL was Enhanced within Skills | 2 | .617 | .471 | .626 |
| AWL is important for EFL | 2 | .225 | .269 | .765 |
| AWL is important for PostGrad | 2 | .383 | .482 | .619 |
| I've Been Taught AWL in Class | 2 | .261 | .267 | .766 |
| AWL I learned Was Enough | 2 | 2.440 | 1.511 | .227 |
| I Wished I Had More AWL | 2 | .857 | .803 | .452 |
| I Had AWL out Class Set By Teacher | 2 | 2.357 | 1.675 | .194 |
| I learned AWL out Class By My Self | 2 | .619 | .477 | .622 |
| **Questions:** | | | | |
| I Used Text Books to study AWL | 2 | .271 | 1.161 | .319 |
| I Had AWL in Class | 2 | .221 | 1.137 | .326 |
| Had Homework By Teacher | 2 | .081 | .648 | .526 |
| I Self Learned the AWL | 2 | .929 | 2.988 | .057 |
| No AWL study This Term | 2 | .042 | .832 | .439 |
| Error | 164 | | | |

### Level of study effect:

| | | Value Label | N |
|---|---|---|---|
| Level | 0 | None | 3 |
| | 1 | UnderGrad | 84 |
| | 2 | Master | 80 |
| | 3 | ResearchDegree | 1 |

| Dependent Variable | df | M-Square | F | Sig. |
|---|---|---|---|---|
| **Rates**: | | | | |
| Learned AWL | 3 | 2.083 | 1.629 | .185 |
| Learn And Can Write Speak AWL | 3 | 4.291 | 4.606 | .004 |
| Course Focused On AWL | 3 | .441 | .301 | .825 |
| AWL was Enhanced within Skills | 3 | 3.319 | 2.371 | .072 |
| AWL is important for EFL | 3 | 1.695 | 1.291 | .279 |
| AWL is important for PostGrad | 3 | 1.761 | 1.131 | .338 |
| I've Been Taught AWL in Class | 3 | 53.170 | 40.937 | .000 |
| AWL I learned Was Enough | 3 | 14.123 | 9.096 | .000 |
| I Wished I Had More AWL | 3 | 1.820 | 1.255 | .292 |
| I Had AWL out Class Set By Teacher | 3 | 1.000 | .810 | .490 |
| I learned AWL out Class By My Self | 3 | .485 | .396 | .756 |
| **Questions**: | | | | |
| I Used Text Books to study AWL | 3 | .481 | 1.963 | .122 |
| I Had AWL in Class | 3 | .176 | .857 | .465 |
| Had Homework By Teacher | 3 | 3.215 | 15.887 | .000 |
| I Self Learned the AWL | 3 | .183 | .634 | .594 |
| No AWL study This Term | 3 | .032 | 1.367 | .255 |
| Error | 164 | | | |

## EFL years of study effect

| | | Value Label | N |
|---|---|---|---|
| EFL years of study | 0 | 0-5Yrs | 4 |
| | 1 | 6-7Yrs | 35 |
| | 2 | 8-9Yrs | 36 |
| | 3 | 10-11Yrs | 17 |
| | 4 | over 12Yrs | 76 |

| Dependent Variable | df | M-Square | F | Sig. |
|---|---|---|---|---|
| **Rates**: | | | | |
| Learned AWL | 4 | 2.672 | 2.122 | .080 |
| Learn And Can Write Speak AWL | 4 | 2.023 | 2.093 | .084 |
| Course Focused On AWL | 4 | .966 | .661 | .620 |
| AWL was Enhanced within Skills | 4 | 4.259 | 3.120 | .017 |
| AWL is important for EFL | 4 | 1.558 | 1.185 | .319 |
| AWL is important for PostGrad | 4 | 1.785 | 1.148 | .336 |
| I've Been Taught AWL in Class | 4 | .639 | .282 | .890 |
| AWL I learned Was Enough | 4 | 1.432 | .801 | .526 |
| I Wished I Had More AWL | 4 | .855 | .581 | .677 |
| I Had AWL out Class Set By Teacher | 4 | .410 | .328 | .859 |
| I learned AWL out Class By My Self | 4 | 1.146 | .946 | .439 |
| **Questions:** | | | | |
| I Used Text Books to study AWL | 4 | .136 | .542 | .705 |
| I Had AWL in Class | 4 | .284 | 1.395 | .238 |
| Had Homework By Teacher | 4 | .072 | .276 | .893 |
| I Self Learned the AWL | 4 | .372 | 1.304 | .271 |
| No AWL study This Term | 4 | .074 | 3.353 | .011 |
| Error | 164 | | | |

## Test scores effect:

| | | Value Label | N |
|---|---|---|---|
| Test Scores | 2.0 | 5.5 | 1 |
| | 4.5 | 4.5 | 1 |
| | 5.0 | 7.0 | 11 |
| | 5.5 | 5.5 | 51 |
| | 6.0 | 6.0 | 17 |
| | 6.5 | 6.5 | 6 |
| | 7.0 | 7.0 | 1 |

| Dependent Variable | df | M-Square | F | Sig. |
|---|---|---|---|---|
| **Rates:** | | | | |
| Learned AWL | 6 | 2.594 | 1.775 | .115 |
| Learn And Can Write Speak AWL | 6 | 2.677 | 2.568 | .025 |
| Course Focused On AWL | 6 | 1.266 | .739 | .620 |
| AWL was Enhanced within Skills | 6 | 2.491 | 1.785 | .112 |
| AWL is important for EFL | 6 | 1.382 | 1.022 | .417 |
| AWL is important for PostGrad | 6 | 2.238 | 1.269 | .281 |
| I've Been Taught AWL in Class | 6 | 4.749 | 2.154 | .056 |
| AWL I learned Was Enough | 6 | 5.446 | 3.078 | .009 |
| I Wished I Had More AWL | 6 | 2.194 | 1.375 | .235 |
| I Had AWL out Class Set By Teacher | 6 | 2.410 | 2.203 | .051 |
| I learned AWL out Class By My Self | 6 | 2.085 | 1.739 | .122 |
| **Questions**: | | | | |
| I Used Text Books to study AWL | 6 | .181 | .702 | .649 |
| I Had AWL in Class | 6 | .184 | .967 | .453 |
| Had Homework By Teacher | 6 | .454 | 1.735 | .123 |
| I Self Learned the AWL | 6 | .256 | 1.014 | .422 |
| No AWL study This Term | 6 | .005 | .232 | .965 |
| Error | 164 | | | |